LARGE DEVIATIONS AND APPLICATIONS

Srinivasa R.S. VARADHAN

TABLE DES MATIERES

S.R.S.VARADHAN : "LARGE DEVIATIONS AND APPLICATIONS"

Large Deviations and Applications

S.R.S. Varadhan
New York University
Courant Institute of Mathematical Sciences
251 Mercer Street
New York, NY 10012

Section 1. Introduction, Definitions and Examples.

There are many situations arising in analysis, physics and other areas where solutions to certain problems are naturally expressed in terms of function space integrals or expectations of certain functionals with respect to specific stochastic processes. This representation of the solution can be used for many purposes. First of all it can be used to prove the existence of the solution, and it can also be used to establish some of the qualitative properties of the solution. One may also be able to perform some Monte Carlo simulations to evaluate the integral.

However, we will take a different point of view in these lectures. Often there are one or more parameters in the problem and we have a situation where the functional to be integrated as well as the probability measure to be used in the integration may depend on these parameters. Asymptotic evaluation of these integrals when the parameter becomes large or small is rather useful. Whereas in a single integral contributions come from the entire range of integration it is quite conceivable that as the parameters approach their extreme values the integration process becomes singular in the sense that the major contribution to the integral comes from a set whose measure is becoming extremely small. The principle of large deviation is the art of determining how small the probabilities of these rare events really are. It is then used to identify where the major contribution to the integral comes from and leads to a precise estimation of the integral itself.

Let us examine this by means of a simple example. Let $x_1, x_2, \cdots, x_n, \cdots$ be a sequence of independent positive random variable with a common distribution α. We will assume for simplicity that $1 \le x_i \le 2$ with probability 1. Let $\xi_n = x_1 x_2 \cdots x_n$. Then $\log \xi_n = \sum_1^n \log x_i$ and $\lim_{n\to\infty} \frac{1}{n} \log \xi_n = \int_1^2 \log x \, d\alpha(x)$, i.e. with probability nearly one we expect $\xi_n \sim a^n$ where $\log a = \int_1^2 \log x \, d\alpha(x)$. On the other hand $E\xi_n = (Ex_1)^n = (\int_1^2 x \, d\alpha(x))^n$. By Jensen's inequality $Ex_1 \ge \exp(E\log x_1)$ and if α is nondegenerate the inequality is strict. Therefore the contribution to $E\xi_n$ from

typical sequences which grow like a^n does not account for the growth in $E\xi_n$. Where does the contribution come from? We might try to analyze what the probability is for ξ_n to grow like λ^n. Of course unless $1 \le \lambda \le 2$ this probability is zero. Unless $\lambda = \exp(E\log x_1)$ this probability goes to zero. It turns out that the probability goes to zero exponentially rapidly, i.e. like $[\rho(\lambda)]^n$ where $\rho(\lambda)$ can be explicitly determined as a function of λ. The quantity $\rho(\lambda)$ is one if $\lambda = \exp(E\log x_1)$, $0 \le \rho(\lambda) \le 1$ for all values of λ and $\rho(\lambda) = 0$ unless $0 \le \lambda \le 1$. The contribution to the integral $E\xi_n$ from those ξ_n which are like λ^n is then $\lambda^n[\rho(\lambda)]^n$. The maximum contribution comes from that value of λ_0 such that

$$\lambda_0\rho(\lambda_0) = \sup_{0\le\lambda\le 1} \lambda\rho(\lambda)$$

and $Ex_1 = \lambda_0\rho(\lambda_0)$. The principle of large deviation tells us which values of ξ_n contribute to $E\xi_n$ as $n \to \infty$.

A more complicated example is provided by considering a Markov chain on a finite state space X. For $x,y \in X$ let $\pi(x,y)$ be the probability of transition from state x to state y in a single step. We will assume for simplicity that $\pi(x,y) > 0$ for all $x,y \in X$. Under this condition we have a unique invariant set of probabilities $\alpha(x)$, $x \in X$ and we have the usual set of ergodic theorems on the long term behavior of irreducible finite state space Markov Chains. Let us take a function $V: X \to R$ and consider $E_x\exp[V(X_1) + \cdot\cdot V(X_n)] = J_n(x)$ where x is the starting point or X_0 and $X_1,\cdots X_n$ are the states of the Markov Chain at times 1 through n. If we take a frequency count $f_n(y)$ of the number of visits to the state y by the string $X_1,\cdots,X_n$ then

$$\begin{aligned} J_n(x) &= E_x\exp[\sum_y V(y)f_n(y)] \\ &= E_x\exp[n \sum_y V(y) \frac{f_n(y)}{n}] . \end{aligned}$$

By the ergodic theorem $\frac{f_n(y)}{n}$ is very close to $\alpha(y)$ for all states y with a very high probability. If $\beta(y);\ y \in X$ is any other probability vector on the state space $P_x[\frac{f_n(y)}{n} \sim \beta(y)$ for all $y \in X]$ is likely to be very small if $\beta \ne \alpha$. In fact the probability is exponentially small with a rate $\exp[-nI(\beta)]$ which is independent of the starting point to within the exponential constant. The contribution to $I_n(x)$

from the strings $X_1,\cdots,X_n$ whose relative proportion of visits to the states is close to $\beta(\cdot)$ is $\exp[n\Sigma V(y)\beta(y) - nI(\beta)]$ so that one expects

$$(1.1) \qquad \lim_{n\to\infty} \frac{1}{n} \log J_n(x) = \sup_{\beta} [\Sigma V(y)\beta(y) - I(\beta)]$$

where the supremum is taken over all probability vectors β on X. $I(\alpha)$ of course is zero and $I(\beta) > 0$ for $\beta \neq \alpha$. We can of course evaluate $J_n(x) = \langle\delta_x,(\pi e^V)^n e\rangle$ where e^V is the diagonal matrix with $\exp\{V(x)\}$ on the diagonal place corresponding x. Here e is the vector of units and δ_x is the vector with 1 at x and 0 everywhere else. Frobenius theory of positive matrices evaluates

$$(1.2) \qquad \lim_{n\to\infty} \frac{1}{n} \log J_n(x) = \log \lambda(V)$$

where $\lambda(V)$ is the spectral radius of the positive matrix $\{\pi(x,y)\exp[V(y)]\}_{x,y\in X}$. So we obtain a relation

$$(1.3) \qquad \log \lambda(V) = \sup_{\beta} [\Sigma V(y)\beta(y) - I(\beta)] .$$

It turns out that $I(\beta)$ and $\log \lambda(V)$ are convex functions of β and V respectively and we have the dual relationship

$$(1.4) \qquad I(\beta) = \sup_{V} [\Sigma V(y)\beta(y) - \log \lambda(V)] .$$

Technically one proves the exponential rates of decay for sets of interest by variants of the above formula.

We will conclude this lecture by providing a formal definition of what we mean by the principle of large deviation. The definition is formulated in terms of exponential rates of decay.

Let X be a complete separable metric space and let P_n be a sequence of probability measures on the Borel σ-field of X. A function $I(x)$ from $X \to R$ is called a rate function if

(i) $0 \leq I(x) \leq \infty$

(ii) $I(x)$ is lower semi-continuous on X, and

(iii) The level sets $A_\ell = \{x: I(x) \leq \ell\}$ are compact sets in X.

The sequence $\{P_n\}$ is said to obey a principle of large deviation with rate function $I(\cdot)$ if

a) For every closed set $C \subset X$

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n(C) \leq - \inf_{x\in C} I(x)$$

and

b) For every open set $G \subset X$

$$\liminf_{n\to\infty} \frac{1}{n} \log P_n(G) \geq -\inf_{x\varepsilon G} I(x) .$$

It follows then that if A is a Borel set with

$$\inf_{x\varepsilon A^0} I(x) = \inf_{x\varepsilon \bar{A}} I(x)$$

then

$$\lim_{n\to\infty} \frac{1}{n} \log P_n(A) = -\inf_{x\varepsilon A} I(x) .$$

Here A^0 and $\bar{A}$ are respectively the interior and closure of A.

Sometimes one runs into a situation where the rate function satisfies only (i) and (ii) and only a weak form of the large deviation principle with

a) replaced by

a') For every compact set $K \subset X$

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n(K) \leq -\inf_{x\varepsilon K} I(x)$$

holds.

b) Stays the same. In such situations it is difficult to aggregate the local estimates provided by the rate function $I(\cdot)$ into a global estimate. Of course if the space X is compact then there is no difference between the two.

Section 2. Basic General Facts.

In this lecture we will establish some of the basic implications that follow from the principle of large deviations.

Let P_n satisfy the principle of large devitions on a complete separable metric space with a rate function $I(\cdot)$. Let $F(\cdot)$ be a bounded continuous function on X. Then

Theorem 2.1

$$\lim_{n\to\infty} \frac{1}{n} \log \int \exp[nF(x)]dP_n(x) = \sup_{x} [F(x) - I(x)] .$$

Proof: Given any $\delta > 0$ we can find a finite number $C_1, C_2, \cdots, C_N$ of closed sets such that the oscillation of $F(x)$ on C_j is less than δ for every j and such that $C_1, \cdots, C_N$ cover X. Then

$$\int_X \exp[nF(x)]dP_n \le \sum_{j=1}^{N} \int_{C_j} \exp[nF(x)]dP_n$$

$$\le \sum_{j=1}^{N} \exp[n \sup_{x\varepsilon C_j} F(x)]\cdot P_n(C_j)$$

$$\le \sum_{j=1}^{N} \exp[n \inf_{x\varepsilon C_j} F(x) + n\delta]\cdot P_n(C_j) .$$

Therefore

$$\limsup_{n\to\infty} \frac{1}{n} \log \int_X \exp[n\, F(x)]dP_n$$

$$\le \sup_{1\le j\le N} [\inf_{x\varepsilon C_j} F(x) + \delta - \inf_{x\varepsilon C_j} I(x)]$$

$$\le \sup_{1\le j\le N} [\sup_{x\varepsilon C_j} \{F(x) - I(x)\} + \delta\}$$

$$= \sup_{x\varepsilon X} [F(x) - I(x)] + \delta .$$

Since $\delta > 0$ is arbitrary we have

$$\text{(2.1)} \qquad \limsup_{n\to\infty} \frac{1}{n} \log \int_X \exp[n\, F(x)]dP_n \le \sup_{x\varepsilon X} [F(x) - I(x)]$$

On the other hand if $x \varepsilon X$ is arbitrary and U is a neighborhood of x with $F(y) \ge F(x) - \varepsilon$ for $y \varepsilon U$ then

$$\int_X \exp[n\, F(y)]dP_n \ge \int_U \exp[n\, F(y)]dP_n$$

$$\ge \exp[n\, (F(x) - \varepsilon)]P_n(U) .$$

Therefore

$$\liminf_{n\to\infty} \frac{1}{n} \int_X \exp[n\, F(y)]dP_n$$

$$\ge F(x) - \varepsilon - \inf_{y\varepsilon U} I(y)$$

$$\ge F(x) - I(x) - \varepsilon .$$

Since $x \varepsilon X$ and $\varepsilon > 0$ are arbitrary we obtain

$$\text{(2.2)} \qquad \liminf_{n\to\infty} \frac{1}{n} \log \int_X \exp[n\, F(y)]dP_n(y) \ge \sup_{x} [F(x) - I(x)]$$

(2.1) and (2.2) establish the theorem.

Sometimes one has to use a more complex version of theorem 2.1. We will state and prove the part of this version relevant for upper bounds.

<u>Theorem 2.2</u>. Let $F_n(x)$ be a sequence of nonnegative functions such that for some lower semicontinuous non negative function $F(x)$

$$\liminf_{n\to\infty} F_n(x_n) \geq F(x)$$

for every sequence $x_n \to x$, and every $x \in X$. Then

$$\limsup_{n\to\infty} \frac{1}{n} \log \int_X \exp[-n\, F_n(x)]dP_n$$
$$\leq - \inf_{x\in X} [F(x) + I(x)] \ .$$

Proof: Let $\ell = \inf_{x\in X}[F(x) + I(x)]$. For any $\delta > 0$ and $x \in X$ there is a neighborhood $U_{\delta,x}$ of x such that

$$\inf_{y\in \overline{U}_{\delta,x}} I(y) \geq I(x) - \delta$$

and

$$\liminf_{n\to\infty} \inf_{y\in U_{\delta,x}} F(y) \geq F(x) - \delta$$

Therefore as $n \to \infty$

$$\int_{U_{\delta,x}} \exp[-n\, F_n(y)]dP_n \leq P_n(\overline{U}_{\delta,x})\exp[-n(F(x) - \delta) + o(n)] \ .$$

Since $U_{\delta,x}$ is a closed set,

$$\limsup_{n\to\infty} \frac{1}{n} \log \int_{U_{\delta,x}} \exp[-nF_n(y)]dP_n$$
$$\leq -[F(x) - \delta] - [I(x) - \delta]$$
$$\leq -\inf_{y\in U_{\delta,x}} [F(x) + I(x)] + 2\delta \ .$$

If K is any compact set in X, then a finite union of $U_{\delta,x}$ as x varies will cover K. Let us call this finite union U_δ. We have

$$\limsup_{n\to\infty} \frac{1}{n} \log \int_{U_\delta} \exp[-nF_n(y)]dP_n$$
$$\leq -\inf_{y\in U_\delta} [F(x) + I(x)] + 2\delta$$
$$\leq -\ell + 2\delta \ .$$

Let us pick $K = \{x\colon I(x) \leq k\}$ where $k \gg \ell$. Then

$$\limsup_{n\to\infty} \frac{1}{n} \log \int_{X-U_\delta} \exp[-n\, F_n(y)]dP_n$$
$$\leq \limsup_{n\to\infty} \frac{1}{n} \log P_n[X - U_\delta]$$
$$\leq -\inf_{x\in U_\delta} I(x) \leq -k \ .$$

Therefore

$$\limsup_{n\to\infty} \frac{1}{n} \log \int_X \exp[-n\, F_n(y)]dP_n$$
$$\leq \text{Max}[-\ell + 2\delta,\ -k]\ .$$

Since $k < \infty$ is arbitrary and $\delta > 0$ is arbitrary we let $k \to \infty$ and $\delta \to 0$ to obtain our theorem.

A situation that comes up often in applications is the following: P_n is a sequence of probability measures on X satisfying a large deivation principle with a rate function $I(\cdot)$. We have a continuous map $F: X \to Y$ into another complete separable metric space. We denote by $Q_n = P_n \circ F^{-1}$, the image of P_n on Y udner F. One can ask if Q_n satisfies a large deviation principle and if so what is the relation of its rate function to $I(\cdot)$.

Theorem 2.3. Q_n satisfies a large deviation principle with a rate function $I'(y)$ given by

$$I'(y) = \inf_{x:F(x)=y} I(x)\ .$$

Proof: It follows from the properties of the rate function $I(\cdot)$ that $I'(\cdot)$ is a rate function on Y. Moreover for any closed set $C \subset Y$

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n} \log Q_n(C) &= \limsup_{n\to\infty} \frac{1}{n} \log P_n(F^{-1}C) \\ &\leq -\inf_{x\in F^{-1}(C)} I(x) \\ &= -\inf_{y\in C} I'(y)\ .\end{aligned}$$

The lower bound for open sets is similar. We will refer to this theorem as the contraction principle.

The theorems in the theory of large deviations are fairly stable under reasonable perturbations; for instance if we assume that P_n satisfies a large deviation principle with rate $I(\cdot)$ and F_n are continuous maps from $X \to Y$ converging uniformly on compact sets of X to F then for the image $Q_n = P_n \circ F_n^{-1}$ we have again a theorem.

Theorem 2.4. $Q_n = P_n \circ F_n^{-1}$ satisfies a large deivation principle with the rate function

$$I'(y) = \inf_{x:F(x)=y} I(x)\ .$$

<u>Proof</u>: Let $A \subset Y$ be a closed. Let $C_n = \{x : F_n(x) \in A\}$ then

$$P_n(C_n) = Q_n(A) .$$

If we let $C = \{x : F(x) \in A\}$, from the uniform convergence of F_n to F on compact sets it follows that given any open set U containing C and any compact $K \subset X$ there is a neighorhood K^δ of K such that

$$C_n \cap K^\delta \subset U \quad \text{for sufficiently large } n .$$

Therefore for n large enough

$$\begin{aligned} P_n(C_n) &\leq P_n(U) + P_n(X - K^\delta) \\ &\leq P_n(\bar{U}) + P_n(X - K^\delta) . \end{aligned}$$

If we take $K = \{x : I(x) \leq \ell\}$ we obtain

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n(C_n) \leq \operatorname{Max}[- \inf_{x\in\bar{U}} I(x) , -\ell] .$$

We let $\ell \to \infty$ and $\bar{U} \downarrow C$ so that

$$\lim_{\bar{U}\downarrow C} \inf_{x\in\bar{U}} I(x) = \inf_{x\in C} I(x)$$

and we obtain our desired result.

Let us now take $G \subset Y$ to be open. Let $y \in G$ be arbitrary and $x \in X$ be such that $F(x) = y$. Since $F_n(x)$ tends to $F(x)$ uniformly on compact sets we can find a neighborhood V of x in X such that $F_n(V) \subset G$ for n large enough. Therefore for sufficiently large n

$$Q_n(G) \geq P_n(V)$$

and

$$\begin{aligned} \liminf_{n\to\infty} \frac{1}{n} \log Q_n(G) &\geq \liminf_{n\to\infty} \frac{1}{n} \log P_n(V) \\ &\geq -I(x) . \end{aligned}$$

Since this is true for every $x \in X$ such that $y = F(x) \in G$ we have our theorem.

<u>Section 3. Large deviations for stationary stochastic processes</u>.

Our lectures deal mainly with large deviations of various ergodic phenomena. Let us take a sequence $X_1, X_2, \cdots X_n, \cdots$ of real valued random variables which form an ergodic stationary sequence in the strict sense. We can extend the process to

nonpositive integers and obtain a stationary process $\{X_j\}$ $-\infty < j < \infty$. The measure P corresponding to such a process is a translation invariant ergodic measure on the doubly infinite sequences of real numbers. For every n we have the random variable $\xi_n = \frac{X_1 + \cdots X_n}{n}$ and this will have a distribution Q_n under P. The ergodic theorem asserts that if $E^P|X_i| < \infty$ then Q_n for large n is close to the degernate distribution at $a = E^P X_i \in R$. Hopefully under suitable susmptions on the underlying process P, Q_n will satisfy a large deviation principle on R with a rate function $h(x)$, $x \in R$. Since the entire probability is getting concentrated at $x = a$ we expect $h(a) = 0$ and $h(x) > 0$ for $x \neq a$.

We can consider a more general situation in which the stationary process $\{X_j\}$ takes values in an arbitrary Polish space X. We can take for our ξ_n the random varible $\xi_n(f) = \frac{f(X_1) + \cdots f(X_n)}{n}$ where f is a bounded continuous function on X. Again we can expect to strengthen the ergodic theorem by establishing a large deviation principle for the distributions Q_n^f of $\xi_n(f)$ under P on R with a rate function $h_f(y)$, $y \in R$. We cn be more ambitious and consider a vector $f_1, \cdots, f_d$ of functions on X and expect for the random vector $\xi_n(f)$ or for its distribution Q_n^f on R^d under P a large deviation principle with a rate function $h_f(y)$, $y \in R^d$.

In fact one should be even more ambitious and consider the random variable

$$\xi_n = \frac{\delta_{X_1} + \cdots \delta_{X_n}}{n}$$

as a map from the space Ω of infinite sequences into the space of probability measure on X. ξ_n is then the empirical distribution of the process based on the first n random variables. The distribution of ξ_n is then a measure Q_n on the spce M of probability measures on X. The ergodic theorem is still valid and asserts that ξ_n converges almost surely (in the sense of weak convergence in the space M) with respect to P to the measure α which is the one dimensional distribution of the random variable X_i under P. We may again expect for Q_n a principle of large deviation to hold in the space M with a rate function $I(\mu)$, $\mu \in M$. Needless to say, we expect $I(\alpha) = 0$ and $I(\mu) > 0$ for $\mu \neq \alpha$. If f is a bounded continuous vector valued function on X (with values in R^d) then

$$\frac{f(X_1) + \cdots f(X_n)}{n} = \int f \, d\xi_n .$$

Therefore the map $\mu \to \int f d\mu$ from $M \to R^d$ maps Q_n onto Q_n^f. By the contraction

principle one may deduce the large deviation principle for Q_n^f from that of Q_n. The relation between the rate functions is provided by Theorem 2.2. Therefore

$$h_f(y) = \inf_{\mu:\int f d\mu = y} I(\mu) .$$

We have so far looked at the ergodic theorems for random variables of the form

$$\xi_n(f) = \frac{f(X_1) + \cdots, f(X_n)}{n} .$$

But the ergodic theorems apply equally well for random variables of the form

$$(3.1) \qquad \xi_n(g) = \frac{g(X_1,X_2) + \cdots g(X_n,X_{n+1})}{n} .$$

Continuing on in the same spirit we might want to look at the map

$$\xi_n^{(2)} = \frac{\delta_{X_1,X_2} + \cdots \delta_{X_n,X_{n+1}}}{n} .$$

from Ω into $M^{(2)}$ the space of measures on $X \times X$ or $X^{(2)}$. We may expect again a large deviation principle with a rate function $I^{(2)}(\beta)$ for $\beta \in M^{(2)}$. There is nothing special about 2 and we may take

$$\xi_n^{(k)} = \frac{\delta_{X_1,\cdots X_K} + \delta_{X_2,\cdots X_{k+1}} + \cdots + \delta_{X_n,\cdots X_{n+k-1}}}{n} .$$

In fact we should abandon all restraint and consider $\xi_n^{(k)}$ for the totality of all possible values of k. This has to be done with a little care. For every sequence

$$\omega = (\cdots x_{-1}, x_0, x_1, \cdots) \in \Omega$$

and every n let us consider the sequence

$$\omega^{(n)} = (\cdots x_1, \cdots, x_n, x_1, \cdots x_n, x_1, \cdots, x_n, \cdots) .$$

Formally the i^{th} coordinate of $\omega^{(n)}$ is given by

$$\omega_i^{(n)} = x_i \qquad \text{if } 1 \le i \le n$$

$$\omega_{i+n}^{(n)} = \omega_i^{(n)} \qquad \text{for all } i \text{ and } n .$$

In other words we keep the chunk of ω from x_1, through x_n and make it periodic outside of period n. If we look at all the periodic sequences of period n in Ω and denote this set by $\Omega^{(n)}$ then the map $\omega \to \omega^{(n)}$ defines a map π_n from $\Omega \to \Omega^{(n)}$. Given any point $\omega^{(n)}$ in $\Omega^{(n)}$, denoting by T the shift in Ω, $\omega^{(n)}, T\omega^{(n)}, \cdots, T^{n-1}\omega^{(n)}$ is a periodic orbit in $\Omega^{(n)}$ and $\frac{\delta_{\omega^{(n)}} + \delta_{T\omega^{(n)}} + \cdots \delta_{T^{n-1}\omega^{(n)}}}{n}$ defines a T invariant measure on $\Omega^{(n)}$. This is of course a stationary stochastic process on $\Omega^{(n)}$ and since

$\Omega^{(n)}$ Ω this is a stationary process on Ω. In this manner for each n and ω we have defined a stationary process

$$R_{n,\omega} = \frac{1}{n}(\delta_{\omega}(n) + \delta_{T\omega}(n) + \cdots + \delta_{T^{n-1}\omega}(n))$$

If $g(x_1,x_2)$ is viewed as a map from $\Omega \to R$ then

$$\int g(x_1,x_2)dR_{n,\omega} = \frac{1}{n}[g(x_1,x_2) + \cdots g(x_{n-1},x_n) + g(x_n,x_1)] .$$

This is not quite the same as what we have in (3.1) but the difference is just one term in n and becomes negligible as $n \to \infty$. The ergodic theorem again tells us that

$$P[\lim_{n\to\infty} R_{n,\omega} = P] = 1 .$$

We might as well expect a large deviation principle for the distribution $\hat{Q}_n$ of $R_{n,\omega}$ under P. Now the state space is the space of all stationary stochastic processes and we expect a rate function H(Q) for $Q \in \mathcal{M}_S$ which is equal to zero only when Q = P. There is of course a natural map from $\mathcal{M}_S \to M$ which assigns to any stationary process its common one dimensional marginal distribution. If we call this map τ

$$\tau R_{n,\omega} = \xi_n = \frac{1}{n}(\delta_{x_1} + \cdots \delta_{x_n})$$

Since map τ is continuous from $\mathcal{M}_S \to M$ the contraction principle applies and we can have a large deviation principle in M if we have on in $\mathcal{M}_S$. The rate functions are of course related by

$$I(\mu) = \inf_{Q:\tau Q=\mu} H(Q) .$$

Of course to actually carry all of this out requires serious assumptions on the nature of the underlying stationary process P. We will, during these lectures, start with the special case of independent random variables or product measure for P. Then, we will look at the case of a Stationary Markov Chain. We will also look at Stationary Gaussian Processes. We will then extend these results to the case of continuous time Markov Process. Towards the end we will look at some applications of the theories developed here.

Section 4. Independent Random Variables.

Throughout these lectures, in all instances the rate functions will have a

close connection with some sort of entropy. It is therefore important for us to spend some time establishing some of the properties of entropy.

Definition: Given any two probability measures β and α on a measure space (X,Σ) we define the entropy of β with respect to α as

$$h(\beta;\alpha) = \sup_{V \in B_0} [\int V(x)d\beta(x) - \log\int e^{V(x)}d\alpha(x)] \tag{4.1}$$

where B_0 is the space of all bounded measurable functions on X.

This definition is the same as relative entropy or Kullback-Liebler information number: This is the content of the following theorem:

Theorem 4.1. The following two statements are equivalent:

a) $h(\beta,\alpha) = \ell < \infty$.

b) $\beta << \alpha$ and if $f(x) = \frac{d\beta}{d\alpha}$

then $f(x)\log f(x)$ is integrable with respect to α and

$$\int f(x)\log f(x)d\alpha(x) = \ell .$$

Proof: Let us first assume that b) holds for some finite ℓ. Then using $xy \leq y\log y + e^{x-1}$ valid for x real and $y > 0$

$$\int V(x)d\beta(x) = \int V(x)f(x)d\alpha(x)$$

$$\leq \int f(x)\log f(x)d\alpha(x) + \frac{1}{e}\int e^{V(x)}d\alpha(x)$$

we can write $V(x) = (V(x) - k) + k$. Then

$$\int V(x)d\beta(x) \leq \int f(x)\log f(x)d\alpha(x) + e^{-(k+1)}\int e^{V(x)}d\alpha(x) + k .$$

We pick $k = \log\int e^{V(x)}d\alpha(x) - 1$. This yields

$$\int V(x)d\beta(x) \leq \ell + \log \int e^{V(x)}d\alpha(x) .$$

Since V is arbitrary we establish a).

Let us now assume that a) is true. First we want to show that $\beta << \alpha$. let A be any set. Take $V(x) = k\chi_A(x)$. We get from a)

$$k\beta(A) \leq \ell + \log(1 - \alpha(A) + \alpha(A)e^k)$$

or

$$\beta(A) \leq \frac{1}{k}[\ell + \log(1 + \alpha(A)(e^k - 1)]$$

$$\leq \inf_{k>0} \frac{1}{k}[\ell + \log(1 + \alpha(A)(e^k - 1))]$$

$= \psi(\ell,\alpha(A))$

where

(4.2) $$\psi(\ell,\delta) = \inf_{k>0} \frac{1}{k}[\ell + \log(1 + \delta(e^k - 1))] .$$

it is clear that $\psi(\ell,\delta) \to 0$ as $\delta \to 0$ for each fixed $\ell < \infty$ so that not only is $\beta << \alpha$ but any time ℓ is controlled such β are uniformly absolutely continuous with respect to α. We now have

$$\sup_{V\varepsilon B_0} \int V(x)f(x)d\alpha(x) \leq \ell + \log \int e^{V(\alpha)}d\alpha(x) .$$

We would like to take $V(x) = \log f(x)$ to otain b). But we do not know that $f(x)\log f(x)$ is integrable with respect to α. In any case we may only take bounded V. We pick

$$V_{\varepsilon,k} = \log[(f\wedge k)\vee\varepsilon] .$$

We let $\varepsilon \to 0$ and then $k \to \infty$. Since flogf is bounded near $f = 0$, letting $\varepsilon \to 0$ is no problem. Finally as $k \to \infty$ we use the monotone convergence theorem to establish our result. The details are left as an exercise.

Sometimes when we are dealing with a polish space X and its Borel sets for Σ, it is convenient to have the following lemma:

Lemma 4.2. If X is any Polish space the supremum in definition (4.1) can be taken over the class of bounded continuous functions and we will still have the same supremum as over all bounded measurable functions:

Proof: a trivial application of Lusin's theorem.

Suppose we have two probability measures β and α on (X,Σ) and a sub σ-field $\Sigma_0 \subset \Sigma$. Then we may just look at β and α on Σ_0 and restricting V to be bounded and measurable with respect to Σ_0 we obtain what we might call $h_{\Sigma_0}(\beta;\alpha)$. In other words the relative entropy is also with respect to a specified σ-field which may only be a sub σ-field. Obviously if $\Sigma_1 \subset \Sigma_2$ then $h_{\Sigma_1}(\beta;\alpha) \leq h_{\Sigma_2}(\beta;\alpha)$. We want to interpret the difference again as an entropy. Let us suppose that β and α possess regular conditional probabilities β_ω and α_ω given the σ-field Σ_1. Then

Lemma 4.3. We have the following identity

$$h_{\Sigma_2}(\beta;\alpha) = h_{\Sigma_1}(\beta,\alpha) + E^\beta h_{\Sigma_2}(\beta_\omega,\alpha_\omega) .$$

Proof: We can assume without loss of generality that $h_{\Sigma_1}(\beta,\alpha) < \infty$. Otherwise $h_{\Sigma_2}(\beta,\alpha) \geq h_{\Sigma_1}(\beta,\alpha) = \infty$ and the identity is valid because both sides are infinite. If $h_{\Sigma_1}(\beta,\alpha) < \infty$ then $\beta << \alpha$ on Σ_1 an therefore α_ω is defined not only almost everywhere with respect to α, but β as well. Therefore the second term on the right makes sense.

For any V bounded and Σ_2 measurable

$$\begin{aligned}
E^{\beta}[V(x)] &= E^{\beta}E^{\beta_\omega}V(x) \\
&\leq E^{\beta}[\log E^{\alpha_\omega}e^{V(x)}] + E^{\beta}h_{\Sigma_2}(\beta_\omega,\alpha_\omega) \\
&\leq h_{\Sigma_1}(\beta,\alpha) + \log E^{\alpha}E^{\alpha_\omega}e^{V(x)} + E^{\beta}h_{\Sigma_2}(\beta_\omega,\alpha_\omega) \\
&= \log E^{\alpha}e^{V(x)} + h_{\Sigma_1}(\beta,\alpha) + E^{\beta}h_{\Sigma_2}(\beta_\omega,\alpha_\omega) .
\end{aligned}$$

This proves one half of the lemma: As for the other half we note that if $\beta < \alpha$ on Σ_2 then

$$\frac{d\beta}{d\alpha}\Big|\Sigma_2 = \frac{d\beta}{d\alpha}\Big|\Sigma_1 \cdot \frac{d\beta_\omega}{d\alpha_\omega}\Big|\Sigma_2 .$$

Let us now suppose that we have a product measure P on $\Omega = \prod_{-\infty}^{\infty} X_i$ with marginal distribution α on each X_i which are copies of X. Let Q be any statinary process on Ω. We denote by F_m^n the σ-field generated by the coordinates x_i of $\omega \in \Omega$ for $n \leq i \leq m$. If $m = \infty$ the σ-field is denoted by F^n and if $n = -\infty$ by F_m. We denotely Q_ω the regular conditional probability distribution of x_1 given F_0 under Q. The rate function that will play a role here is

Definition 4.4. $H(Q;P) = E^{Q}h_{F_1^1}(Q_\omega,\alpha)$.

Although the functional $H(Q;P)$ is defined through entropy it has several variational formulae as well and we need them in order to establish some of its properties as well as in the proof of the large deviation principle.

Let A_n be the class of bounded measurable functions on $X^{(n)} = X\times\cdots,\times X$ i.e. functions $F(x_1,\cdots,x_n)$, satisfying the condition

$$\int_X \exp[F(x_1,\cdots,x_n)]d\alpha(x_n) \leq 1 \quad \forall\, x_1,\cdots,x_{n-1} .$$

Theorem 4.5.

$$H(Q;P) = \sup_n \sup_{F\in A_n} E^{Q}(F)] .$$

Moreover we can replace A_n by the set of bounded continuous function C_n of n variables instead and we also have

$$H(Q;P) = \sup_n \sup_{F \varepsilon C_n} E^Q[F] .$$

Proof: Let $F \varepsilon A_n$. Then

$$\begin{aligned} E^Q[F] &= E^Q[F(x_1,\cdots,x_n)] \\ &= E^Q E^Q[F(x_1,\cdots,x_n)|F_{n-1}] \\ &= E^Q \log E^P[e^{F(x_1,\cdots x_{n-1})}|F_{n-1}] + H(Q;P) \\ &\leq H(Q;P) . \end{aligned}$$

To establish the reverse inequality we define $\bar{Q}$ on the σ-field F_1 by making $\bar{Q} = Q$ on F_0 and making the first coordinate independent of the past F_0 and having the distribution α. Then $\bar{Q}_\omega = \alpha$. Then we see that

$$\begin{aligned} H(Q;P) &= h_{F_1}(Q,\bar{Q}) \\ &= \sup_{F \varepsilon F_1} E^Q[F] - \log E^{\bar{Q}}[e^F] \\ &= \sup_n \sup_{F=F(x_{-n},\cdots,x_{-1},x_0,x_1)} E^Q[F] - \log E^{\bar{Q}}[e^F] \\ &\quad \text{(by a variant of Lusin's theorem)} \\ &\leq \sup_n \sup_{F \varepsilon F_1^{-n}} E^Q[F] - E^Q[\log \textstyle\int e^{F(\cdots x_1)} d\alpha(x_1)] \\ &= \sup_n \sup_{F \varepsilon A_n} E^Q[F] . \end{aligned}$$

Another way of calculating H(Q,P) is by the following Theorem.

Theorem 4.6.

$$H(Q;P) = \sup_n \frac{1}{n} \sup_{F \varepsilon D_n} E^Q[F]$$

where

$$D_n = \{F: F = F(x_1,\cdots,x_n) \; ; \; \int e^{F(x_1,\cdots,x_n)} dP \leq 1\} .$$

Proof: Let us call $F(x_1,\cdots,x_n)$ by $F_n(x_1,\cdots,x_n)$ and define successively

$$F_k(x_1,\cdots,x_k) = \log \int e^{F_{k+1}(x_1,\cdots,x_{k+1})} d\alpha(x_{k+1}) .$$

Then we verify that $F_0 \leq 0$ and

$$\int e^{F_{k+1}(x_1,\cdots,x_{k+1}) - F_k(x_1,\cdots,x_k)} d\alpha(x_{k+1}) \leq 1 .$$

Therefore by Theorem 4.5

$$E^Q F_{k+1} - E^Q F_k \leq H(Q,P)$$

adding over $k = 0,1,\cdots,n-1$ and using $F_0 \leq 0$ we have

$$E^Q F_n = E^Q F \leq nH(Q;P) \ .$$

To prove the converse we note that by definition

$$\sup_{F \in D_n} E^Q[F] = h_{F_n^1}(Q,P) \ .$$

We compute

$$h_{F_n^1}(Q;P) - h_{F_{n-1}^1}(Q,p)$$
$$= h_{F_1^{2-n}}(Q;P) - h_{F_1^{3-n}}(Q;P) \text{ (by stationarity)}$$
$$= h_{F_1^1}(Q_\omega^{-(n-2)};\alpha) \ .$$

As $n \to \infty$ the above term has a limit inferior of at least $h_{F_1^1}(Q_\omega,\alpha) = H(Q;P)$. This establishes the theorem:

We are now ready to prove our upper and lower bounds for establishing the large deviation principle. We have the measure $\hat{Q}_n$ which is the distribution of $R_{n,\omega}$ under P. We divide the theorem into many lemmas.

Theorem 4.7. For any closed set $C \subset M$

$$\limsup_{n\to\infty} \frac{1}{n} \log \hat{Q}_n(C) \leq -\inf_{Q \in C} H(Q,P) \tag{4.3}$$

and for $G \subset M$, open

$$\liminf_{n\to\infty} \frac{1}{n} \log \hat{Q}_n(G) \geq -\inf_{Q\in G} H(Q;P) \ . \tag{4.4}$$

Proof: Let us denote for any Borel set A

$$J(A) = \limsup_{n\to\infty} \frac{1}{n} \log \hat{Q}_n(A) \ . \tag{4.5}$$

We will establish several properties of $J(A)$ as lemmas leading up to the proof of (4.3). The lower bound (4.4) will be dealt with later.

Lemma 4.8. Let $F(x_1,\cdots,x_k) \in C_k$. Then for $n \geq 1$

$$E^P\{\exp[\tfrac{1}{k}(F(x_1,\cdots,x_k) + F(x_2,\cdots,x_{k+1}) + \cdots F(x_n,\cdots,x_{n+k-1}))\} \leq 1 \ .$$

Proof: We write

$$F(x_1,\cdots,x_k) + \cdots F(x_n,\cdots x_{n+k-1})$$

$$= G_1(x_1,x_2,\cdots) + G_2(x_1,x_2,\cdots) + \cdots G_k(x_1,x_2,\cdots)$$

where

$$G_1 = F(x_1,\cdots,x_k) + F(x_{k+1},\cdots,x_{2k}) + \cdots$$

$$G_2 = F(x_2,\cdots x_{k+1}) + F(x_{k+2},\cdots,x_{2k+1}) + \cdots$$

$$G_k = F(x_k,\cdots,x_{2k-1}) + F(x_{2k},\cdots,x_{3k-1}) + \cdots .$$

Then

$$E^P\{\exp[\frac{1}{k}(F(x_1,\cdots,x_k) + \cdots \quad F(x_n,\cdots,x_{n+k-1}))]\}$$

$$= E^P\{\exp[\frac{1}{k}(G_1 + \cdots + G_k)]\}$$

$$\leq E^P[\frac{1}{k}(e^{G_1} + \cdots + e^{G_k})]$$

$$\leq \frac{1}{k}\sum_1^k E^P[e^{G_i}] \leq 1 .$$

<u>Lemma 4.9</u>. For any Borel set $A \subset M$

$$J(A) \leq -\sup_k \frac{1}{k} \sup_{F\varepsilon C_k} \inf_{Q\varepsilon A} E^Q\{F\} .$$

<u>Proof</u>:

$$|\frac{1}{n}[F(x_1,x_2,x_k) + \cdots + F(x_n,\cdots,x_{n+k-1})] - E^{R_{n,\omega}}\{F\}| \leq \frac{k-1}{n} ||F||$$

where $||F||$ is the sup norm on F.

Therefore from lemma 4.8

$$E^P[\exp[\frac{1}{k}(F(x_1,\cdots,x_k) + \cdots F(x_n,\cdots,x_{n+k-1}))]] \leq C$$

or

$$E^{\hat{Q}_n}[\exp[\frac{n}{k} \int F dQ]] \leq C$$

or

$$\hat{Q}_n(A) \leq Ce^{-\frac{n}{k} \inf_{Q\varepsilon A} \int F dQ} .$$

Taking logs, dividing by n and taking limsup as $n \to \infty$

$$J(A) \leq -\frac{1}{k} \inf_{Q\varepsilon A} \int F \, dQ .$$

Since $F \varepsilon C_k$ and k are arbitrary we have our result.

<u>Lemma 4.10</u>. Let $K \subset M$ be compact and let $\varepsilon > 0$ be given. Then there exists an open set G_ε in M such that $K \subset G_\varepsilon$ and

$$J(G_\varepsilon) \leq - \inf_{Q\varepsilon K} H(Q;P) + \varepsilon .$$

Proof: Let $\varepsilon > 0$ and k be given. For each $Q \varepsilon K$ there exists an integer $k(Q)$ and $F_Q \varepsilon C_{k(Q)}$ such that

$$\frac{1}{k(Q)} \int F_Q dQ \geq H(Q,P) - \varepsilon/2 .$$

Since $\int F_Q dQ$ is a continuous linear functional of Q, there is a neighborhood N_Q of Q such that for $Q' \varepsilon N_Q$

$$\frac{1}{k(Q)} \int F_Q dQ' \geq H(Q,P) - \varepsilon .$$

Therefore from lemma 4.9

$$J(N_Q) \leq - H(Q,P) + \varepsilon .$$

N_Q as Q varies over K is an open covering of K and let $N_{Q_1},\cdots,N_{Q_L}$ be a finite subcover. Denoting by G_ε such a finite subcover clearly G_ε K and

$$\begin{aligned} J(G_\varepsilon) &\leq - \min_{1\leq j\leq L} J(N_{Q_j}) \\ &\leq - \min_{1\leq j\leq L} [H(Q_j,P) - \varepsilon] \\ &\leq - \min_{1\leq j\leq L} H(Q_j;P) + \varepsilon \\ &\leq - \inf_{Q\varepsilon K} H(Q;P) + \varepsilon . \end{aligned}$$

Lemma 4.11. Given any $\ell < \infty$ there exists a compact set $K_\ell \subset M$ such that

$$J(K_\ell^c) \leq - \ell .$$

Proof: For each $Q \varepsilon M$ let us denote by q the marginal distribution of x_0. Then if B_ℓ M is a compact set of probability measures on X. Then

$$K_\ell = \{Q : q \varepsilon B_\ell\} \text{ is compact in } M .$$

We need therefore only estimate

$$\hat{Q}_n\{Q : q \varepsilon B_\ell^c\}$$

or

$$P\{\omega : \frac{\delta_{x_1} + \cdots \delta_{x_n}}{n} \varepsilon B_\ell^c\} .$$

From Prohorov's theorem we can take

$$B_\ell = \{q : q(D_j^\ell) \geq 1 - \frac{1}{j} \quad \text{for } j = 2,\cdots .\}$$

where $D_j^\ell \subset X$ are compact subsets of X for each j and ℓ. Therefore

$$P\{\omega: \frac{\delta_{x_1} + \cdots \delta_{x_n}}{n} \varepsilon B_\ell^c\}$$
$$\leq \sum_{j=2}^{\infty} P\{\omega: \frac{\chi_{D_j^\ell,c}(x_1) + \cdots \chi_{D_j^\ell,c}(x_n)}{n} \geq \frac{1}{j}]$$
$$\leq \sum_{j=2}^{\infty} 2^n e^{-n\ell j} \quad \text{(by Lemma 4.12 with } \theta = \ell j^2)$$

if we pick D_j^ℓ such that $\alpha(D_j^\ell) \geq 1 - e^{-\ell j^2}$ which is possible by Prohorov's theorem. If we assume $\ell \geq 1$, then the lemma follows.

Lemma 4.12. Let $x_1,\cdots,x_n$ be n independent random variables with values 0 or 1 with probability ε and $1 - \varepsilon$. Then for any $\theta > 0$

$$\text{Prob}[\frac{x_1+\cdots,x_n}{n} \geq \delta] \leq e^{-n\theta\delta}(\varepsilon e^\theta + (1 - \varepsilon))^n .$$

Proof: Apply Tchebechev's inequality to

$$E[e^{x_1+\cdots,x_n}] = (\varepsilon e^\theta + 1 - \varepsilon)^n .$$

Proof of (4.3). Given any closed set C we pick an $\ell < \infty$ and the comapct set K_ℓ of lemma 4.11. Then K_ℓ C is compact and

$$J(C) \leq - \text{Min}[J(K_\ell \cap C), J(K_\ell^c)]$$
$$\leq - [\inf_{Q \varepsilon C} H(Q;P),\ell] .$$

Letting $\ell \to \infty$ we obtain our result. We now work on the lower bound:

Let us denote by $\psi(\omega)$ the Radon-Nikodym derivative of Q_ω with respect to α on the σ-field F_1^1. Then $\psi(\omega)$ can be thought of as a measurable function on F_1 and $E^Q \log\psi(\omega) = H(Q;P)$. Moreover if we denote by T the shift on the space of sequences then $\frac{dQ_\omega}{dP}$ on F_n^1 is $\exp[\psi(\omega) + \psi(T\omega) + \cdots\psi(T_\omega^{n-1})]$.

Lemma 4.13. Let Q be any ergodic element in M. Then for any neighborhood N of Q in M

$$\liminf_{n\to\infty} \frac{1}{n} \log \hat{Q}_n[N] \geq - H(Q;P) .$$

Proof: Assume $H(Q;P) < \infty$.

$$\hat{Q}_n(N) = P\{R_{n,\omega} \varepsilon N\}$$

$$= \int_{R_{n,\omega}\varepsilon N} dP$$

$$\geq \int_{R_{n,\omega}\varepsilon N} \frac{dP}{dQ}\, dQ$$

$$= \int_{R_{n,\omega}\varepsilon N} e^{-(\log\psi(\omega)+\cdots\log\psi(T^{n-1}\omega))}\, dQ$$

$$\geq \int_{R_{n,\omega}\varepsilon N \,\cap\, E_\delta} e^{-n(H(Q,P)+\delta)}\, dQ$$

where $E_\delta = \{\omega:\frac{1}{n}\sum_{j=1}^{n} \log \psi(T_\omega^j) \leq H(Q;P) + \delta\}$. By ergodic theorem $Q\{\omega:R_{n,\omega} \varepsilon N\}$ as well as $Q(E_\delta)$ tend to 1 as $n \to \infty$ and we are done:

<u>Lemma 4.14</u>. If $Q \varepsilon M$ is arbitrary then Q has an integral representation

$$Q = \int_{M_e} Q'\pi_Q(dQ')$$

over the ergodic measures and

$$H(Q,P) = \int_{M_e} H(Q',P)\Pi_Q(dQ') .$$

<u>Proof</u>: From standard results in ergodic theory we know that the integral representation is valid. Moreover the regular conditional probability has a version Θ_ω such that

$$\Theta_\omega = Q_\omega \text{ a e } \omega\ Q \quad \forall \varepsilon M .$$

Therefore

$$H(Q;P) = \int h_{F_1^1}(\Theta_\omega,\alpha)dQ$$

clearly satisfies the lemma:

Now we prove the lower bound (4.4). If $Q \varepsilon M$ is arbitrary we can approximate it by a finite linear combination of ergodic ones such that $Q \sim \Sigma\pi_j Q_j$ and $H(Q;P) \sim \Sigma\pi_j HJ(Q_j,P)$. We can therefore assume without loss of generality that $Q = \Sigma\pi_j Q_j$ with Q_j ergodic. Let n be given and define $n_j = \pi_j n$. For a given $\omega = \omega_1$ let $\omega_j = T^{n_1+\cdots n_{j-1}}\omega_1$. Then $R_{n,\omega} \sim \Sigma\pi_j R_{n_j,\omega_j}$ since the only difference is the periodization at the end. Since the topoloogy on M is essentially convergence of finite dimensional distributions for a given finite dimensional range the effect of the periodization goes to zero as $n \to \infty$. Therefore given a neighborhood N of Q, there are neighborhoods N_j of Q_j such that

$$R_{n_j,\omega_j} \varepsilon N_j \text{ for } \forall j \Rightarrow R_{n,\omega} \varepsilon N .$$

Therefore

$$P[R_{n,\omega} \in N] \geq \Pi\, P[R_{n_j,\omega_j} \in N_j] \ .$$

Taking logs, dividing by n and taking limint as $n \to \infty$

$$\varliminf_{n\to\infty} \frac{1}{n} \hat{Q}_n(N) \geq - \Sigma\pi_j H(Q_j,P) \geq - H(Q) - \varepsilon$$

and we are done: We want to conclude this section by stating some properties of various entropy functions:

Lemma 4.15. For fixed α, $h(\beta;\alpha)$ is a lower semicontinuous convex function of β in the weak topology.

Proof:

$$h(\beta;\alpha) = \sup_{F\epsilon C(X)} [\int Fd\beta - \log\int e^F d\alpha] \ .$$

The properties are now obvious:

Lemma 4.16. For fixed P, a product measure based on α, $H(Q;P)$ is lower semicontinuous in Q:

Proof:

$$H(Q,P) = \sup_k \sup_{F\epsilon C_k} [\int FdQ]$$

and the lemma is obvious:

Lemma 4.17. For fixed β

$$\inf_{Q:q=\beta} H(Q;P) = h(\beta;\alpha)$$

and the inf is attained at the product measure:

Proof: It is clear that for the product measure Q_β

$$H(Q_\beta;P) = h(\beta;\alpha) \ .$$

C_k contains functions $e^{F(x_1)+\cdots F(x_k)}$ with $\int e^{F(x)} d\alpha(x) \leq 1$. Therefore for any Q, $H(Q;P) \geq h(q;\alpha)$.

Lemma 4.18. For any $\ell < \infty$ and any α

$$\{q:h(q;\alpha) \leq \ell\} \text{ is compact in } M \ .$$

Proof: From inequality (4.2) if $\alpha(A) < \delta$ then $q(A) < \eta$ where $\eta = \eta(\delta,\ell) \to 0$ as $\delta \to 0$ for each ℓ. Since α is tight by Prohorov's theorem $\{q:h(q,\alpha) \leq \ell\}$ is uniformly tight and hence compact.

Lemma 4.19. For any ℓ, $\{Q:H(Q,P) \leq \ell\}$ is compact in M.

Proof: As Q varies over our set q varies over a set contained in $\{q:h(q;\alpha) \leq \ell\}$ and is therefore conditionally compact. Therefore so is the set of measures Q.

Section 5. Markov Chains

In this section we will assume the base measure P to be based on a Markov chain rather than a product measure. One difference however is that the P measure will be defined only on F^0 which is all that is needed. Moreover instead of a single P we have a family P_{x_0} depending on the starting point x_0 at time zero. The transition probability is $\pi(x,dy)$. We make the following assumption on $\pi(x,dy)$.

Hypothesis 1.

$\pi(x,dy)$ has the Feller property or for any bounded continuous function $f(\cdot)$ on X,

$$(\pi f)(x) = \int f(y)\pi(x,dy)$$

is bounded and continuous on X.

The random processes $R_{n,\omega}$ are defined as before and instead of $\hat{Q}_n$ we now have $\hat{Q}_{n,x_0}$ depending on the starting point x_0 as well. We will describe the results in this case and indicate in the proof only modifications needed in the earlier proof for the independent case:

Definition 5.1. Given π and $Q \in M$ we define

$$H(Q;\pi) = E^Q\{h_{F_1^1}(Q_\omega,\pi(x_0,\cdot))\} .$$

Here x_0 is thought of as a function of ω and then the relative entropy of Q_ω and $\pi(x_0,\cdot)$ is calculated on the σ-field F_1^1 corresponding to x_1. The answer that depends on ω is averaged with respect to Q.

In theorem 4.5 we should modify A_n so that

$$A_n = \{F:F = F(x_1,\cdots,x_n) \text{ and } \int_X \exp[F(x_1,\cdots,x_n)]\pi(x_{n-1},dx_n) \leq 1 \quad \forall\ x_1,\cdots x_{n-1}\}$$

C_n is defined accordingly:

Then we have

Theorem 5.2.

$$H(Q;\pi) = \sup_n \sup_{F\in A_n} E^Q[F]$$

$$= \sup_n \sup_{F \epsilon C_n} E^Q[F] .$$

Proof:

Proceeds in a manner identical to theorem 4.5 with minor obvious modifications. We next define D_n as

$$D_n = \{F:F = F(x_1,\cdots,x_n) \text{ and } \int e^{F(x_1,\cdots,x_n)} dP_{x_0} \le 1 \quad \forall x_0\} .$$

Then we have the analog of theorem 4.6.

Theorem 5.3.

$$H(Q,P) = \sup_k \frac{1}{k} \sup_{F \epsilon D_k} E^Q[F] .$$

Proof: We define for given $F = F(x_1,\cdots,x_n)$, $F_k = F_k(x_1,\cdots,x_k)$ successively for $k < n$ by

$$F_k(x_1,\cdots,x_k) = \log \int e^{F_{k+1}(x_1,\cdots,x_{k+1})} \pi(x_k,dx_{k+1}) .$$

Then $E^Q[F_{k+1} - F_k] \le H(Q,\pi)$ by theorem 5.2. The proof is completed as before. Now we start establishing the large deviation principle.

First we define for any $x_0 \in X$

$$(5.1) \qquad J_{x_0}(A) = \limsup_{n\to\infty} \frac{1}{n} \log \hat{Q}_{n,x_0}(A) .$$

We then have

Lemma 5.4. For any compact set $K \subset M$ and any $\varepsilon > 0$ there exists a neighborhood G_ε K such that

$$J_{x_0}(G_\varepsilon) \le - \inf_{Q \epsilon K} H(Q;\pi) + \varepsilon .$$

Proof:

Identical to lemmas 4.8, 4.9 and 4.10.

The main difference is only at this point. In order to go from compact K to closed C we need to make a strong positive recurrence assumption on the transition probability $\pi(x,dy)$.

Hypothesis 2

Let us suppose that there are functions $U(x)$ and $V(x)$ on X with the following properties:

a) $U(x) \ge 1$ for all x and $(\pi U)(x)$ is bounded on compact subsets of X.

b) $V(x) = \log U(x) - \log(\pi U)(x)$ is bounded below (away from $-\infty$) and for any ℓ $\{x:V(x) \leq \ell\}$ is a totally bounded subset of X.

Under hypothesis 2 we can establish the analog of lemma 4.11.

Lemma 5.5. Given any $\ell < \infty$, there is a compact set K_ℓ M such that

$$J_{x_0}(K_\ell^c) \leq -\ell \quad \text{for every } x_0 \in X .$$

Proof: The proof of lemma 5.5 will depend on lemma 5.6 and will follow the lines of the proof of lemma 4.11.

Lemma 5.6. Given any ℓ and j there is a compact set D_j^ℓ X such that

$$P_{x_0}\{\omega:\frac{1}{n}[\chi_{D_0^\ell,c}(x_1) + \cdots \chi_{D_j^\ell,c}(x_n)] > \frac{1}{j}\} \leq C^n e^{-\ell nj}$$

for all j and n. Here C is some fixed constant.

Proof: From hypothesis 5.5 we have

$$E^{P_{x_0}}\{e^{V(x_p)+\cdots V(x_n)}U(x_{n+1})\} = \pi U(x_0) .$$

Since $U \geq 1$ and $\pi U(n_0) < \infty$ we have

$$E^{P_{x_0}}\{e^{V(n_1) + \cdots V(x_n)}\} \leq C .$$

If we take $D_j^\ell = \{x:V(x) \leq \lambda\}$ for some λ then

$$V(x_1) + \cdots V(x_n) \geq \lambda \sum_{r=1}^{n} \chi_{D_j^\ell,c}(x_r)]] - nC_1 .$$

where $-C_1$ is a lowerbound for $V(\cdot)$. Therefore

$$E^{P_{x_0}}[\exp[\lambda \sum_{r=1}^{n} \chi_{D_j^\ell,c}(x_r)]] \leq C^n .$$

Therefore $P_{x_0}[\frac{1}{n} \sum_{r=1}^{n} \chi_{D_j^\ell,c}(x_r) \geq \frac{1}{j}] \leq C^n e^{-\frac{\lambda n}{j}}$.

If we choose $\lambda = \ell j^2$ in D_j^ℓ we have our estimate. If we therefore have hypothesis 1 and 2 we can get the upper bound part of the large deviation principle. Moreover one can check through the proof that all estimates are valid uniformly provided x_0 varies over a compact subset of X. Now we start working on the lower bound. We need another hypothesis.

Hypothesis 3. The transition probability $\pi(x,dy)$ has a density $\pi(x,y)$ with respect to a reference measure α such that

a) $\pi(x,y) > 0$ a e α for each $x \in X$ and

b) the map $x \to \pi(x,\cdot)$ is continuous as a map of X into $L_1(\alpha)$.

Theorem 5.7. Let $Q \in M$ be ergodic then for any open N containing Q

$$\lim_{n\to\infty} \frac{1}{n} \log P_{x_0}[R_{n,\omega} \in N \text{ and } x_n \in K_2] \geq -H(Q;\pi)$$

where K_2 is any compact set in X with $\alpha(K_2) > 0$; and the limit is uniform for x_0 varying over any compact set in X.

Proof: Let us pick K_3 such that $q(K_3) \geq \frac{1}{2}$, where q is the marginal of Q. Denoting by

$$\frac{dQ_\omega}{dP_{x_0(\omega)}\big|F_1^1} = \psi(\omega)$$

where $x_0(\omega)$ is the coordinate of ω corresponding to zero we have

$$\frac{dQ_\omega}{dP_{x_0(\omega)}\big|F_1^1} = \exp\left[\sum_{j=0}^{n-1} \log \psi(T^j\omega)\right] .$$

Therefore if we take the set

$$E_{N,n} = \{\omega\colon R_{n,\omega} \in N \text{ and } x_n(\omega) \in K_3\}$$

then

$$P_{x_0(\omega)}(E_{N,n}) \geq \int_{E_{N,n}} \exp\left[-\sum_{j=0}^{n-1} \log \psi(T^j\omega)\right] dQ_\omega$$

$$\geq e^{-[H(Q,P)+\delta]n} Q_\omega\left[E_{N,n} \cap \{\omega\colon \frac{1}{n}\sum_{1}^{n-1} \log \psi(T^j\omega) \leq H(Q;P) + \delta\}\right] .$$

Let us denote by $\phi(n,x)$ the quantity

$$\phi(n,x) = P_x(E_{N,n})e^{[H(Q;P)+\delta]n} \wedge 1 .$$

Then

$$\phi(n,x_0(\omega)) \geq Q_\omega[E_{N,n} \cap D_{n,\delta}]$$

where

$$D_{n,\delta} = \{\omega\colon \frac{1}{n}\sum_{1}^{n-1} \log \psi(T^j\omega) \leq H(Q;P) + \delta\} .$$

Taking expectations with respect to Q we have

$$\int \phi(n,x)dq(x) \geq Q[E_{N,n} \cap D_{n,\delta}] \tag{5.2}$$

$$\to q(K_3) \geq \frac{1}{2} \quad \text{as } n \to \infty .$$

Therefore

$$\liminf_{n\to\infty} \int \phi(n,x)dq(x) \geq \frac{1}{2} .$$

We can find a smaller neighborhood $N_1 \subset N$ such that if $R_{n-2,T\omega} \in N_1$ then $R_{n,\omega} \in N_1$ for n sufficiently large. Therefore for large n

$$\begin{aligned} &P_{x_0}(R_{n,\omega} \in N, x_n \in K_2) \\ &\quad \geq P_{x_0}(R_{n-2,\omega} \in N,\ x_{n-1} \in K_3,\ x_n \in K_2) \\ &\quad \geq \left(\int P_x(E_{N_1,n-2})\pi(x_0,dx)\right)\inf_{x\in K_3} \pi(x,K_2)\ . \end{aligned}$$

From our assumptions the last factor is strictly positive. We denote by θ some lower bound for it. Then

$$\begin{aligned} &P_{x_0}(R_{n,\omega} \in N,\ x_n \in K_2) \\ &\quad \geq \theta\ e^{-[H(Q;P)+\delta]n} \int\phi(n,x)\pi(x_0,dx)\ . \end{aligned}$$

It is now an elementary exercise that our assumptions and (5.2) imply that

$$\liminf_{n\to\infty} \int\phi(n,x)\pi(x_0,dx) > 0$$

and in fact uniformly over compact sets of starting points x_0. We finally have

Theorem 5.8. For C closed in M and G open in M

$$\limsup_{n\to\infty} \frac{1}{n} \log \hat{Q}_{n,x_0}(C) \leq - \inf_{Q\in C} H(Q;P)$$
$$\liminf_{n\to\infty} \frac{1}{n} \log \hat{Q}_{n,x_0}(G) \geq - \inf_{Q\in G} H(Q,P)\ .$$

Proof: All that remains is to pass from ergodic Q to non-ergodic Q. This is carried out exactly like the independent case. Instead of independence we make in each one of time periods the process to have its $R_{n_j,\omega}$ closet to Q_j and end up in a compact set K_2. Since we can afford to take the infimum over the starting point in $x_0 \in K_2$ at the next step it is almost the same as independence.

We also have the results, which are analogs of 4.15 through 4.19.

Lemma 5.9. For each fixed π, $H(Q;\pi)$ is lower semicontinuous in $Q \in M$.

Proof:

$$H(Q;\pi) = \sup_k \sup_{F\in C_k} \left[\int F dQ\right]$$

and the Feller property ensures that the normalization procedure that defined F_k inductively leave the $\{C_k\}$ invariant. Since the functionals on the right are continuous we have lower semi-continuity of $H(Q;\pi)$.

Lemma 5.10. For any $\beta \in M$

$$\inf_{Q:q=\beta} H(Q;P) = \inf_{\lambda\in M_\beta^{(2)}} h_{X^{(2)}}(\lambda;\lambda_0)$$

where λ, λ_0 are probability measures on $X\times X$ and $\lambda_0(dx,dy) = \beta(dx)\pi(x,dy)$. $M_\beta^{(2)}$ consists of all $\lambda \in M^{(2)}$ such that the marginals of both components are β.

Proof: Starting from λ we can construct a unique Markov chain (stationary) whose two dimensional distribution is λ at two consecutive time points. For such a Markov chain Q_λ one can compute

$$H(Q_\lambda;P) = h_{X^{(2)}}(\lambda;\lambda_0) .$$

If Q is not Markov then the $\bar{Q}$ associated to the two dimensional marginal of Q is always Markov and

$$H(\bar{Q};P) \leq H(Q;P) .$$

Lemma 5.11. For any $\beta \in M$

$$\inf_{\lambda\in M_\beta^{(2)}} H_{X^{(2)}}(\lambda,\lambda_0) = \sup_u \int \log \frac{u(x)}{(\pi u)(x)}\, d\beta(x)$$

where the supremum is taken over all bounded uniformly positive measurable functions:

Proof: For the proof we do not need the Feller condition on π so by a standard result on Polish spaces we might as well assume that X is compact. Then we may restrict u to bounded continuous functions.

Suppose for some u, $\int \log \frac{u(x)}{(\pi u)(x)}\, d\beta(x) = \ell$. Then,

$$\int \log \frac{u(y)}{(\pi u)(x)}\, \lambda(dx,dy) = \ell \quad \text{because } \lambda \in M_\beta^{(2)}$$

on the other hand

$$\log \int \frac{u(y)}{(\pi u)(x)}\, \lambda_0(dx,dy) = \log \int \frac{u(y)}{(\pi u)(x)}\, d\beta(x)\pi(x,dy)$$
$$= \log 1$$
$$= 0 .$$

By definition of $H_{X^{(2)}}(\lambda;\lambda_0)$ we have now

$$\lambda \in M_\beta^{(2)} \Rightarrow h_{X^{(2)}}(\lambda;\lambda_0) \geq \ell$$

and we have the easy half. For the other half what we have to show is that if

$$\inf_{\lambda\in M_\beta^{(2)}} h_{X^{(2)}}(\lambda;\lambda_0) \geq \ell \tag{5.3}$$

then we have to produce a continuous u for which $\int \log \frac{u(x)}{(\pi u)(x)}\, d\beta(x) \geq \lambda - \varepsilon$, where $\varepsilon > 0$ is given. this requires the use of the minimax theorem. From (5.3) we have

$$\inf_{\lambda \varepsilon M_\beta^{(2)}} \sup_V [\int V(x,y)\lambda(dx,dy) - \log \int e^{V(x,y)}\lambda_0(dx,dy)] \geq \ell .$$

By standard minimax theorem we can interchange sup and inf so that

$$\sup_V \inf_{\lambda \varepsilon M_\beta^{(2)}} [\int V(x,y)\lambda(dx,dy) - \log \int e^{V(x,y)}\lambda_0(dx,dy)] \geq \ell .$$

In other words given $\varepsilon > 0$ there is a V such that

$$\inf_{\lambda \varepsilon M_\beta^{(2)}} \int V(x,y)\lambda(dx,dy) \geq \ell + \log \int e^{V(x,y)}\lambda_0(dx,dy) - \varepsilon .$$

By normalization we may assume the existence of a V such that

$$(5.4) \qquad \ell + \log \int e^{V(x,y)}\lambda_0(dx,dy) \leq \varepsilon$$

and

$$(5.5) \qquad \int V(x,y)\lambda(dx,dy) \geq 0 \qquad \forall\ \lambda \ \varepsilon\ M_\beta^{(2)} .$$

We may rewrite (5.5) as

$$\inf_\lambda \sup_{\phi,\psi} [\int V(x,y)\lambda(dx,dy) + \int[\phi(y) + \psi(x)]\lambda(dx,dy) - \int[\phi(x) + \psi(x)]\beta(dx)] \geq 0$$

because the sup is 0 if $\lambda \ \varepsilon\ M_\beta^{(2)}$ and ∞ otherwise. Again by minimax theorem (5.5) implies

$$(5.6) \ \sup_{\phi,\psi} \inf_\lambda [\int V(x,y)\lambda(dx,dy) + \int[\phi(y) + \psi(x)]\lambda(dx,dy) - \int[\phi(x) + \psi(x)]d\beta(x)] \geq 0$$

which means that given any $\varepsilon > 0$, there is pair ϕ,ψ such that (again by normalization)

$$(5.7) \qquad \int \phi d\beta = \int \psi d\beta = 0$$

and

$$(5.8) \qquad V(x,y) \geq \phi(x) + \psi(y) - \varepsilon \quad \forall \ \text{x and y}$$

(5.4) and (5.8) yield

$$(5.9) \qquad \ell + \log \int e^{\phi(x)+\psi(y)}\beta(dx)\pi(x,dy) \leq 2\varepsilon .$$

If we call $e^\psi = u$ then (5.9) is the same as

$$(5.10) \qquad \log \int e^{\phi(x)}(\pi u)(x)\beta(dx) \leq 2\varepsilon - \ell .$$

By Jensen's inquality we get

$$\int \phi(x)\beta(dx) + \int \log(\pi u)(x)\beta(dx) \leq 2\varepsilon - \ell$$

since $\log u = \psi$ from (5.7) we obtain

$$\int \log \frac{u(x)}{(\pi u)(x)}\, \beta(dx) \geq \ell - 2\varepsilon$$

and we are done.

If we define

$$I_\pi(\beta) = \sup_u \int \log \frac{u(x)}{(\pi u)(x)}\, \beta(dx) = \inf_{Q:q=\beta} H(Q;\pi)$$

then

Lemma 5.12. $I_\pi(\beta)$ is lower semi-continuous and convex. Under Hypothesis 2 the set $\{\beta : I_\pi(\beta) \leq \ell\}$ is compact in M. And under the same hypothesis $\{Q : H(Q;\pi) \leq \ell\}$ is compact in M.

Proof: By standard truncation we will have

$$\int V(x)d\beta \leq \ell$$

where $V(x)$ is the function of hypothesis 2). By Tchebyshev bounds we obtain the first part of our lemma. The second part follows trivially from the first part.

Section 6. Stationary Gaussian Process

For P we take a stationary Gaussian process with mean 0 and covariance

$$E\{x_n x_{n+k}\} = \rho_k = \frac{1}{2\pi}\int_0^{2\pi} e^{ik\theta} f(\theta)d\theta$$

where $f(\theta)$ is a continuous nonnegative function with $f(0) = f(2\pi)$. We assume that the process is nondeterministic so that $\int_0^{2\pi} \log f(\theta)d\theta$ is greater than $-\infty$.

We construct $R_{n,\omega}$ and $\hat{Q}_n$ and we aim to show that a large deviation principle is valid for $\hat{Q}_n$ with a rate function

$$(6.1) \quad H(Q;f) = E^Q\{\int_{-\infty}^{\infty} q(y/\omega)\log q(y/\omega)dy\}$$

$$+ \frac{1}{2}\log 2\pi + \frac{1}{4\pi}\int_0^{2\pi} \frac{dG(\theta)}{f(\theta)} + \frac{1}{4\pi}\int_0^{2\pi} \log f(\theta)d\theta$$

where $dG(\theta)$ is the spectral measure of Q i.e.

$$E^Q x_0 x_k = \frac{1}{2\pi}\int e^{ik\theta}dG(\theta) .$$

We will outline the basic steps involved in the proof of the large deviation principle for $\hat{Q}_n$ with the rate function provided by (6.1).

Step 1. We represent the random process $\{x_n\}$ as a moving average of the form

$$x_k = \sum_{n=-\infty}^{\infty} a_{n-k}\xi_n$$

where ξ_k are independent random variables and

$$\sqrt{f(\theta)} = \sum_{n=-\infty}^{\infty} a_n e^{in\theta} .$$

The sequence $\{a_n\}$ is in $\ell_2(z)$.

Step 2. We approximate a_n by $a_n^N = a_n(1 - \frac{|n|}{N})$ for $|n| \leq N$ and $a_n^N = 0$ otherwise. If we write

$$g_N(\theta) = \sum_{n=-\infty}^{\infty} a_n^N e^{in\theta}$$

then $g_N(\theta) \to \sqrt{f(\theta)}$ uniformly by Fejer's theorem:

Step 3. Let us define a map on $\Omega = \prod_{i\varepsilon z} R$ by

$$(\tau\omega)(k) = \Sigma\, a_{n-k}\omega(n)$$

then $P = P_0\tau^{-1}$ where P_0 is the product measure based on standard Guassians. If we define τ_N by

$$(\tau_N\omega)(k) = \Sigma\, a_{n-k}^N\omega(n)$$

then $P_N = P_0\tau_N^{-1}$ is Gaussian with mean 0 and spectral density

$$f_N(\theta) = |g_N(\theta)|^2 .$$

Step 4. For each N, $R_{n,\tau_N\omega}$ and $R_{n,\omega}\tau_n^{-1}$ are very close, as $n \to \infty$. In fact any difference between them is only due to periodization. They are both random stationary process and the large deviation principle for $R_{n,\omega}\tau_N^{-1}$ implies the large deviation principle for $R_{n,\tau_N\omega}$. Moreover since we have a large deviation principle for $R_{n,\omega}$ when the basic distribution is P_0, we have one for $R_{n,\omega}\tau_N^{-1}$ since $Q \to Q\tau_N^{-1}$ is a continuous map of M into M. The rate function for $R_{n,\tau_N\omega}$ whose distribution we call $\hat{Q}_{n,N}$ is given by

$$\inf_{Q':Q'\tau_N^{-1}=Q} H(Q',1) .$$

Step 5. We calculate

$$\inf_{Q':Q'\tau_N^{-1}=Q} H(Q',1) = H(Q,f_N)$$

$H(Q,f)$ for any f is given by formula (6.1). Step 5 is mainly a calculation.

Step 6.

$$\lim_{N\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log P_0\{d(R_{n,\tau_N\omega}, R_{\tau\omega}) \geq \varepsilon\} = -\infty$$

for every $\varepsilon > 0$. This is again a calculation based on routine estimates for $\tau_N\omega - \tau\omega$.

Step 7.

$$\hat{Q}_n[C] = P[R_{n,\omega} \in C] = P_0[R_{n,\tau\omega} \in C]$$

$$\leq P_0[R_{n,\tau_N\omega} \in \bar{C}^\varepsilon] + P_0[d(R_{n,\tau\omega}, R_{n,\tau\omega}) \geq \varepsilon] .$$

Taking logs dividing by n, taking limsup and then letting $N \to \infty$ first and then $\varepsilon \to 0$, we obtain

$$\limsup_{n\to\infty} \frac{1}{n} \log \hat{Q}_n(C) \leq - \lim_{\varepsilon\to 0} \liminf_{N\to\infty} \inf_{Q\in\bar{C}_\varepsilon} H(Q;f_N)$$

and similarly for G open

$$\liminf_{n\to\infty} \frac{1}{n} \log \hat{Q}_n(G) \geq - \liminf_{N\to\infty} \inf_{Q\in G} H(Q,f_N) .$$

Step 8.

$$\lim_{\varepsilon\to 0} \liminf_{N\to\infty} \inf_{Q\in\bar{C}_\varepsilon} H(Q;f_N) \geq \inf_{Q\in C} H(Q;f)$$

$$\lim_{N\to\infty} \inf_{Q\in G} H(Q,f_N) \leq \inf_{Q\in G} H(Q;f) .$$

These two statements are proved by the explicit formulas for $H(Q;f_N)$ and $H(Q,f)$ and the explicit definition of f_N in terms of f. Finally

Step 9. $H(Q;f)$ is a rate function.

Section 7. Continuous Time Markov Processes

We will now assume that we have a Markov process with transition probabilities $p(t,x,dy)$ on a state space X with the following properties: The state space X is Polish. Moreover:

Hypothesis 1. The semigroup $(T_tf)(x) = \int f(y)p(t,x,dy)$ maps bounded continuous functions $C(X)$ into itself. For any starting point the measure P_x on the space of

trajectories lives on $\Omega_0 = D[0,\infty)$ which is given the topology of Skorohod convergence on finite intervals. The map $x \to P_x$ is continuous.

Hypothesis 2. There exists a sequence u_n of functions in the domain D of the infinitesimal generator of the process with the following properties

a) $u_n(x) \geq 1$

b) $\sup_{x \in K} \sup_n u_n(x) < \infty$ for each compact K X

c) $\lim_{n\to\infty} u_n(x) = u(x)$ exists for each x. If $V_n(x) = -\frac{LU_n(x)}{U_n(x)}$ then

d) $V_n(x) \geq -C$ for some C for all n,x.

e) $V(x) = \lim_{n\to\infty} v_n(x)$ exists

f) for each $\ell < \infty$, $\{x: V(x) \leq \ell\}$ has compact closure in X.

Hypothesis 3. $p(1,x,dy)$ has density $p(1,x,y)$ for every x with respect to a reference measure α on X. Moreover $p(1,x,y) > 0$ a e α for each x. In addition the map $x \to p(1,x,\cdot)$ is continuous as a map of X into $L_1(\alpha)$.

We denote by Ω the Skorohod space $D(-\infty,\infty)$ and by M the space of stationary processes on Ω. M is a Polish space under weak convergence of processes on finite intervals and the projection map $\omega \to \omega(t)$ while not continuous in general is continuous at almost all points with respect to every $Q \in M$. [Q has no fixed points of discontinuity] for each $\omega \in \Omega_0$ we define $R_{t,\omega}$ by the continuous analog of $R_{n,\omega}$. We extend the trajectory $\omega(s)$, $0 \leq s \leq t$ periodically on either side to get a periodic orbit under the shift θ_s of period t and take $R_{t,\omega}$ as the orbital measure. For ech $x \in X$ we have the distribution $\hat{Q}_{t,x}$ of $R_{t,\omega}$ under P_x. We are interested in a large deviation principle for $\hat{Q}_{t,x}$ on M with some rate function $H(Q)$. We will suppress the dependence of $H(Q)$ on $p(t,x,dy)$, which will be a fixed semigroup for our discussion.

The proof follows the discrete case very closely and we outline the proof giving details only where there are new aspects in the proof.

Definition 7.1. Given $Q \in M$ we define

$$H(Q,T) = E^Q h_{F_T^0}(Q_\omega, P_{\omega(0)}) \ .$$

Lemma 7.2.

$H(Q,T) = TH(Q)$ for some $0 \leq H(Q) \leq \infty$.

Proof: One checks by stationarity of Q and $p(t,x,dy)$ that $H(Q,T_1 + T_2) = H(Q,T_1) + H(Q,T_2)$. Since $H(Q,T) \geq 0$ it follows that $H(Q,T)$ is linear in T.

For each T we define A_T by

$$A_T = \{F : F \text{ is } F_T^0 \text{ measurable and } E^{P_x}\{\exp[F(\omega)]\} \leq 1 \quad \forall\, x\}$$

$$C_T = A_T \quad \{F : E^Q\{F\} \text{ is a continuous linear functinal of } Q \text{ in } M\} .$$

We then have

Theorem 7.3.

$$H(Q) = \sup_{T>0} \frac{1}{T} \sup_{F \varepsilon A_T} E^Q\{F\} = \sup_{T>0} \frac{1}{T} \sup_{F \varepsilon C_T} E^Q\{F\} .$$

Proof: Same as theorem 5.3.

We can define

$$B_T = \{F : F \;\varepsilon\; F_1^{-T} \text{ and } E^{P_{\omega(0)}} e^{F(\omega)} \leq 1 \text{ everywhere } \} .$$

In the above definition integration with respect to $E^{P_{\omega(0)}}$ is carried out over F_1^0 only on each fiber of F_0^{-T}. We also have the analog of theorem 5.2 proved in exactly the same manner.

Theorem 7.4

$$H(Q) = \sup_{T>0} \sup_{F \varepsilon B_T} E^Q\{F\} .$$

We now start proving the large deviation principle.

We define for $x_0 \;\varepsilon\; X$ and $A \subset M$

$$J_{x_0}(A) = \limsup_{T\to\infty} \frac{1}{T} \log \hat{Q}_{T,x_0}(A) .$$

We then have

Theorem 7.5. For any compact set $K \subset M$ and any $\varepsilon > 0$ there exists a neighborhood $G_\varepsilon \supset K$ such that

$$J_{x_0}(G_\varepsilon) \leq -\inf_{Q \varepsilon K} H(Q) + \varepsilon .$$

Proof: Identical to the discrete case.

One main difference in the continuous case is that processes whose marginals vary over a compact set is not necessarily from a compact set of processes. We need to control the modulus of continuity as well.

Theorem 7.6. Let A closed in M be such that the family of one dimensional marginals of Q as Q varies over A forms a tight family of measures on X. Then

$$J_{x_0}(A) \leq - \inf_{Q\varepsilon A} H(Q) .$$

Proof: Let us denote by A_M the family of marginals of A. Given any sequence $\varepsilon_n \to 0$, there exists $K_n \subset X$ such that $q(K_n) \geq 1 - \varepsilon_n$ for $q \varepsilon A_M$. Since $x_0 \to P_{x_0}$ is weakly continuous there exists $C_n \subset D[0,1]$ such that C_n is compact there and $P_x(C_n) \geq 1 - \eta_n$ for all $x \varepsilon K_n$. Denoting by $\tilde{C}_n$ the comnplement of C_n it is easily checked that for all $x \varepsilon X$

$$E^{P_x}\{\exp[\nu X_{K_n}(\omega(0))X_{\tilde{C}_n}(\omega)\} \leq 1 + \eta_n(e^\nu - 1) .$$

From the continuous analog of lemma 4.8

$$E^{P_x}\{\exp[\int_0^t \nu X_{K_n}(\omega(s))\, X_{\tilde{C}_n}(\theta_s\omega)]\} \leq \exp[t \log(1 + \eta_n(e^\nu - 1))] .$$

Therefore allowing for an error of $\frac{1}{t}$ for periodization

$$\hat{Q}_{t,x_0}\{A \cap \{Q:Q(\hat{C}_n) \geq \frac{1}{t} + 2\, \varepsilon_n\}$$

$$\leq \exp[t \log[1 + \eta_n(e^\nu - 1)] - \varepsilon_n \nu t] .$$

Pick $\lambda > 0$, $\nu = \lambda n^2$, $\varepsilon_n = \frac{1}{n}$ and $\eta_n = \exp[-\lambda n^2]$. Then

$$\hat{Q}_{t,x_0}\{A \cap \{Q:Q(\tilde{C}_n) \geq \frac{1}{t} + \frac{2}{n}\} \leq e^{t\log 2-\lambda nt} .$$

If we let

$$A_t = \{Q:Q(\tilde{C}_n) \leq \frac{1}{t} + \frac{2}{n} \text{ for all } n \geq 1\}$$

then

$$\hat{Q}_{t,x_0}\{A \cap \tilde{A}_t\} \leq e^{t\log 2}(\frac{e^{-\lambda t}}{1-e^{-\lambda t}}) .$$

Therefore

$$\limsup_{t\to\infty} \frac{1}{t} \log \hat{Q}_{t,x_0}(A \cap \tilde{A}_t) \leq \log 2 - \lambda . \tag{7.1}$$

It is easy to check tht $A_\infty = \cap_t A_t$ is compact in M and if $G \supset A \cap A_\infty$ is open that

$A_t \cap A \subset G$ for t sufficiently large. Theorem 7.5 and 7.1 provide a proof of Theorem 7.6. If we now assume Hypothesis 2) one can obtain easily

$$E^{P_{x_0}}[\exp \int_0^t V(\omega(s))ds] \leq C$$

for every $t \geq 0$. With this estimate the proof of the upper bound for closed sets proceeds exactly like the discrete case.

Lower bound: The only essential difference with the discrete case is the ergodic theorem: If

$$\frac{dQ_\omega}{dP_{\omega(0)}}\Big|F_t^0 = \exp[\psi(t,\omega)]$$

then $\psi(t + s,\omega) = \psi(t,\omega) + \psi(s,\theta_t\omega)$ is an additive functional. To establish the ergodic theorem almost everywhere we need to show

$$E^Q \sup_{0\leq t\leq 1} |\psi(t,\omega)| < \infty .$$

This is the content of the following lemma:

Lemma 7.7. Let α,β be two probability measures on a measurable space (X,F). Let F_t $0 \leq t \leq 1$ be an increasing family of subfields with $F_1 = F$. Let $h(\beta;\alpha) < \infty$ and $\psi(\omega,t) = \log \frac{d\beta}{d\alpha}\Big|F_t$. Then

$$E^\beta[\sup_{0\leq t\leq 1} |\psi(\omega,t)|] < \infty .$$

Proof: From standard martingale inequalities

$$\beta\{\omega: \inf_{0\leq t\leq 1} \psi(\omega,t) \leq -\ell\} = \beta\{\omega: \inf_{0\leq t\leq 1} R(t,\omega) \leq e^{-\ell}\} \leq e^{-\ell} .$$

We therefore only have to show

$$E^\beta[\sup_{0\leq t\leq 1} \psi(\omega,t)] < \infty .$$

But from entropy inequalities it is sufficient to show that

$$E^\alpha[\sup_{0\leq t\leq 1} e^{\psi(\omega,t)}] < \infty$$

or

$$E^\alpha[\sup_{0\leq t\leq 1} R(t,\omega)] < \infty .$$

Since R is a martingale we need only the integrability $R(\log R)^+$ which is of course true because $h(\beta,\alpha) < \infty$. Now the lower bound is completed as before and we have

Theorem 7.8. Under hypothesis (1), (2) and (3) a large deviation principle holds with the rate function $H(Q)$. In particular $H(Q)$ is a rate function:

We can define for $\mu \in M$

$$I(\mu) = -\inf_{\substack{u>0 \\ u\in D}} \int \left(\frac{Lu}{u}\right)(x)\mu(dx) .$$

Assuming that the domain is big enough we have if

$$I_h(\mu) = -\inf_{u>0} \int \log \frac{\Pi_h u}{u} (x)\mu(dx)$$

then

$$I_h(\mu) \leq h\, I(\mu)$$

and

$$\lim_{h\to 0} \frac{1}{h} I_h(\mu) = I(\mu) .$$

We can use this to prove the contraction principle:

Theorem 7.9

$$\inf_{Q:q=\mu} H(Q) = I(\mu) .$$

Proof: If we denote by Q_h the Markov chain at times that are multiples of h, with invariant measure which is obtained in lemmas 5.10 and 5.11 then the entropy of Q_h with respect to the basic π_h Markov chain with initial distribution μ is given by $I_h(\mu)$ per time gap h. Therefore on the h "grid" in the unit interval $h(Q_h,P_{\mu,h}) \leq \frac{1}{h} I_h(\mu) \leq I(\mu)$. We can fill in the gaps of the grid by the conditional distributions of bridges of span h. This leads to the measure P_μ as filled in $P_{\mu,h}$ and some measure Q_h' for filled in Q_h . $h(Q_h',P_\mu) = h(Q_h,P_{\mu,h}) \leq I(\mu)$. It is easy to show from this entropy estimate the tightness of Q_h' and the limit Q is stationary, has marginal μ and $H(Q) \leq I(\mu)$. The other half of the theorem is just as trivial as the discrete case.

Finally if $p(t,x,dy)$ has a symmetric density $p(t,x,y)$ with respect to a reference measure α then

Theorem 7.10. $I(\mu) < \infty$ if and only if $\mu \ll \alpha$ and if $f = \frac{d\mu}{d\alpha}$ then $\sqrt{f} \in L_2(\alpha)$ is in the domain of $(-L)^{1/2}$ in $L_2(\alpha)$.

$$I(\mu) = ||(-L)^{1/2}\sqrt{f}||^2 .$$

Section 8. Application to the Problem of the Wiener Sausage

A problem that comes up in the study of density of states for Schrödinger

operators with certain random potentials near the edge of the energy spectrum is the following:

Let $\beta(t)$ be d-dimensional Brownian motion starting from the origin. Let $\varepsilon > 0$ be fixed. Consider

$$C_t = \{x : x = \beta(s) \text{ for some } 0 \leq s \leq t\} ,$$

$$C_t^\varepsilon = \{ \bigcup_{x\varepsilon C_t} s(x,\varepsilon)\} = \{x : |x - \beta(s)| < \varepsilon \text{ for some } 0 \leq s \leq t\} .$$

C_t is just the image of the Wiener path up to time t, and C_t^ε is the sausage arount it of radius ε.

The problem is to show that

$$\lim_{t\to\infty} \frac{1}{t^{d/(d+2)}} \log E\{e^{-v|C_t^\varepsilon|}\} = -k(d,v) \tag{8.1}$$

exists and is nonzero. Here $|C_t^\varepsilon|$ is the d-dimensional volume of the sausage C_t^ε up to time t. The actual physical problem involves the Brownian motion that is conditioned to return to the origin at time t. But for large t, one can see easily that the difference between the free Brownian motion and the conditional Brownian motion is small enough that the formula (8.1) is unaffected by it. We will study only the free Brownian motion.

We will first carry out a Brownian change of scale so that $t^{d/(d+2)}$ appears naturally

Let us replace $\beta(s)$, $0 \leq s \leq t$, by

$$t^{1/(d+2)}\beta(t^{-2/(d+2)}s) \quad \text{for } 0 \leq s \leq t^{d(d+2)} ,$$

which is again a Brownian motion. Therefore the distribution of $|C_t^\varepsilon|$ is the same as that of $t^{d(d+2)}|C^{\varepsilon t^{-1/(d+2)}}_{d/(d+2)}|$. If we let $\tau = t^{d/(d+2)}$, then

$$E\{e^{-v|C_t^\varepsilon|}\} = E\{\exp[-\tau v|C_\tau^{\varepsilon\tau^{-1/d}}|]| .$$

The problem therefore reduces to showing that

$$\lim_{t\to\infty} \frac{1}{t} \log E\{\exp[-vt|C_t^{\varepsilon t^{-1/d}}|]\} = - k(v,d) \tag{8.2}$$

exists and is nonzero.

A basic fact in what follows is the behavior of

$$P[\beta(s) \varepsilon G \text{ for } 0 \leq s \leq t]$$

where G is a smooth bounded open set containing the origin. If $L(t,\omega)$ is the random measure representing the occupation time of the Brownian motion, i.e., if

$$L(t,\omega)(A) = \frac{1}{t}\int_0^t \chi_A(\beta(s))ds ,$$

then

$$\beta(s) \in G \text{ for } 0 \leq s \leq t \iff \text{supp } L(t, \omega) \subset G .$$

We can therefore estimate

$$P[\beta(s) \in G \text{ for } 0 \leq s \leq t] \leq P[\text{supp } L(t,\omega) \quad \bar{G}] .$$

Since the set $\{\mu:\text{supp } \mu \quad \bar{G}\}$ is a compact set, we can estimate

$$\limsup_{t\to\infty} \frac{1}{t} \log P[\beta(s) \in G \text{ for } 0 \leq s \leq t]$$

$$\leq - \inf_{\mu:\mu(\bar{G})=1} I(\mu)$$

$$\leq - \inf_{\substack{f=0 \text{ off } \bar{G} \\ f\geq 0, \int f dx=1}} [\frac{1}{8}\int \frac{|\nabla f|^2}{f} dx]$$

$$= - \inf_{\substack{g=0 \text{ off } \bar{G} \\ \int g^2 dx=1}} [\frac{1}{2}\int |\nabla g|^2 dx] = \lambda(G) .$$

The lower estimate

$$\liminf_{t\to\infty} \frac{1}{t} \log P[\beta(s) \in G \text{ for } 0 \leq s \leq t] \geq -\lambda(G)$$

is also easy to derive. See [1] for details. We can combine the two halves in the form of a lemma:

Lemma 8.1. _For nice open sets_ G

$$\lim_{t\to\infty} \frac{1}{t} \log P[\beta(s) \in G \text{ for } 0 \leq s \leq t] = -\lambda(G) .$$

Lower bound. Let G be a bounded open set containing the origin with smooth boundary. Let $\lambda(G)$ be the first eigenvalue of $-\frac{1}{2}\Delta$ in G with Dirichlet boundary conditions. Then according to Lemma 8.1

$$P[\beta(s) \in G \text{ for } 0 \leq s \leq t] = \exp[-\lambda(G) + o(t)] .$$

If $\beta(s) \in G$ for $0 \leq s \leq t$, then

$$|C\xi_t^{-1/d}| \leq |G| + o(1) \quad \text{as } t \to \infty .$$

Therefore

$$\liminf_{t\to\infty} \frac{1}{t} \log E\{\exp[-v|C_t^{\varepsilon t^{-1/d}}|]\} \geq (v|G| + \lambda(G)) .$$

Taking the infimum over all such sets G, we find

$$\liminf_{t\to\infty} \frac{1}{t} \log E\{\exp[-v|C_t^{\varepsilon t^{-1/d}}|]\} \geq -k(v,d)$$

where

$$k(v,d) = \inf_G [v|G| + \lambda(G)] . \tag{8.3}$$

We have proved

Theorem 8.2.

$$\liminf_{t\to\infty} \frac{1}{t} \log E\{\exp[-v|C_t^{\varepsilon t^{-1/d}}|]\} \geq -k(v,d) .$$

We now turn to the

Derivation of the upper bound. Let us replace R^d by the d-dimensional torus T_ℓ^d and consider Brownian motion on the torus. We will denote by E_ℓ expectation with respect to the Brownian motion on the torus of size ℓ. Since any set in R^d projected to T_ℓ^d has volume no larger than the original volume,

$$E\{\exp[-v|C_t^{\varepsilon t^{-1/d}}|]\} \leq E_\ell\{\exp[-v|C_t^{\varepsilon t^{-1/d}}|]\} .$$

We will show that

$$\lim_{\ell\to\infty} \limsup_{t\to\infty} \frac{1}{t} \log E_\ell\{\exp[-v|C_t^{\varepsilon t^{-1/d}}|]\} \leq -k(v,d) .$$

Upper bounds on the torus. Let $\phi(x)$ be a function with support $\{x:|x| \leq \varepsilon\}$ which is nonnegative and has $\int\phi(x)dx = 1$. Then, if $\phi_t(x) = t\phi(xt^{1/d})$, then

$$C_t^{\varepsilon t^{-1/d}} = \{x:L_t * \phi_t > 0\}$$

where L_t is the occupation distribution and * denotes convolution. We will denote by f_t,

$$f_t = L_t * \phi_t ,$$

the mollified local time. The problem we have reduces to two lemmas:

Lemma 8.3.

$$\limsup_{t\to\infty} \frac{1}{t} \log E_\ell\{\exp[-vt|x:f_t > 0|]\} \leq - \inf_f [I(f) + v|x;f > 0|]$$
$$= \inf_G [v|G| + \lambda_\ell(G)] ,$$

and

Lemma 8.4.

$$\lim_{\ell\to\infty} \inf_G [v|G| + \lambda_\ell(G)] = \inf_G [v|G| + \lambda(G)] .$$

Of the two lemmas, the second is a standard approximation lemma involving truncation methods. We will not carry out the proof, but will only refer to [1]. We will sketch a proof of Lemma 8.3. If we consider the L_1 topology for densities on T_ℓ^d, then $|x:f > 0|$ is a lower semicontinuous functional of the density f. In view of Theorem 2.2, it is sufficient to prove the large deviation principle for f_t in the L_1 topology with a rate function $I(f)$.

Lemma 8.5. Let ψ be any mollifier, i.e., a smooth probability density. Then

$$\limsup_{t\to\infty} \frac{1}{t} \log P[f_t * \psi \in C] \leq - \inf_{f:f*\psi\in C} I(f)$$

for any C closed in L_1.

Proof: The map $f \to f * \psi$ is continuous from M with weak topology to L_1 with norm topology. So the large deviation principle in the weak topology for L_t, which implies the large deviation principle in the weak topology for f_t, is converted into a large deviation principle for $f_t*\psi$ in the norm topology of L_1. Theorem 2.3 provides the precise proof. □

We now state without proof Lemma 8.6. We will then state and prove Lemma 8.7, which will imply Lemma 8.3 and our main result. Finally we will prove Lemma 8.6.

Lemma 8.6.

$$\limsup_{t\to\infty} \frac{1}{t} \log P[||f_t * \psi - f_t|| \geq \rho] \leq - k_\rho(\psi)$$

where $k_\rho(\psi) \to \infty$ as $\psi \to \delta_0$ for each $\rho > 0$.

Proof: The proof will be given after the proof of Lemma 8.7.

Lemma 8.7. The large deviation principle holds for f_t with the rate function $I(f)$ in the space L_1 with norm topology.

Proof: Upper bound.

$$P[f_t \in C] \leq P[f_t * \psi \in \bar{C}^\rho] + P[|f_t - f_t * \psi| > \rho] .$$

Therefore

$$\limsup_{t\to\infty} \frac{1}{t} \log P[f_t \in C] \leq - \inf[\inf_{f:f*\psi\in C^\rho} I(f), k_\rho(\psi)] .$$

Letting $\psi \to \delta_0$ and $\rho \to 0$, we get

$$\lim_{\psi\to\delta_0} \inf_{f:f^*\psi\delta\bar{C}^\rho T\rho} I(f) = \inf_{f\epsilon\bar{C}^\rho T\rho} I(f) \text{ and } \lim_{\rho\to 0} \inf_{f\epsilon\bar{C}^\rho} I(f) = \inf_{f\epsilon C} I(f) ,$$

provided C is closed in L_1.

Lower bound. For an open set G around f,

$$P[f_t \epsilon G] \geq P[f_t * \psi \epsilon G_1] - P[||f_t - f_t*\psi|| \geq \rho]$$

where G_1 is a smaller open set around f such that the sphere around G_1 of radius ρ is contained in G. The result is again obvious from Lemma 8.6. □

From Lemma 8.7 we obtain Lemma 8.3 by an application of Thereom 2.2. If we now combine it with the lower bound, i.e. Theorem 8.2, and take Lemma 8.4 for granted, then we have

Theorem 8.8.

$$\lim_{t\to\infty} \frac{1}{t} \log E\{\exp[-v|C_t^{\epsilon t^{-1/d}}|]\} = -k(v,d)$$

where k(v,d) is given by (8.3).

We now turn to the

Proof of Lemma 8.6.

$$\begin{aligned} ||f_t * \psi - f_t|| &= \sup_{|g|\leq 1} |\int (f_t * \psi - f_t(x)g(x)dx| \\ &= \sup_{|g|\leq 1} |\int (L_t * \phi_t * \psi - L_t * \phi_t)(x)g(x)dx| \\ &= \sup_{|g|\leq 1} [\int h_t(x)L_t(dx)] \end{aligned}$$

where

$$h_t(x) = (g * \phi_t * \psi - g * \phi_t)(x) .$$

(We have assumed that ϕ_t and ψ are symmetric.)

The map $g \to \theta$ defined by $\theta = g*\phi_t$ is a compact map of the unit ball $|g| \leq 1$. Therefore for any $\rho > 0$ we can find a finite number $N = N(t,\rho)$ of functions $\theta_1,\cdots,\theta_N$ such that the image of the unit ball is covered by spheres around $\theta_1,\cdots,\theta_N$ of radius $\rho/2$. We can assume that $\theta_1,\cdots,\theta_N$ are all bounded by 1 as well. Then

$$||f_t * \psi - f_t|| \leq \frac{\rho}{4} + \sup_{1\leq i\leq N} [\int (\theta_i(x) - (\theta_i * \psi)(x))L_t(dx)] ,$$

$$P[||f_t * \psi - f_t|| \geq \rho] \leq N \sup_{1\leq i\leq N} P[\int \chi_i(x)L_t(dx) \geq \frac{\rho}{2}]$$

$$\leq N \sup_{1\leq i\leq N} P[\int_0^t \chi_i(\beta(s))ds \geq \frac{\rho}{2} t]$$

$$\leq Ne^{-z\rho t/2}E_\ell\{\exp[z \int_0^t \chi_i(\beta)(s))ds]]$$

where $\chi_i(x) = \theta_i(x) - (\theta_i*\psi)(x)$.

One can show that for any χ with $|\chi| \leq 2$

$$E_\ell\{\exp[z \int_0^t \chi(\beta(s))ds]\} \leq C_z \exp[t\lambda_\ell(z\chi)]$$

where $\lambda_\ell(z\chi)$ is the largest eigenvalue of

$$\frac{1}{2} \Delta + z\chi \quad \text{on } T_\ell^d .$$

If, for each $\rho > 0$, $N(t,\rho) \leq \exp[D_\rho t]$ for some D_ρ, then

$$\frac{1}{t} \log P[||f_t * \psi - f_t|| \geq \rho] \leq D_\rho - \frac{z\rho}{2} + \sup_{\chi:\chi=\theta-\theta*\psi, |\theta|\leq 1} \lambda(z\chi) .$$

One verifies that sup $\lambda(z\chi) \to 0$ as $\psi \to \delta_0$ for each $z > 0$. Therefore

$$\limsup_{\psi\to\delta_0} \limsup_{t\to\infty} \frac{1}{t} \log P[||f_t * \psi - f_t|| \geq \rho] \leq D_\rho - \frac{z\rho}{2} ,$$

and by letting $z \to \infty$ we will obtain our lemma. We no need only the estimation of $N(\rho,t)$ to complete the proof of Lemma 8.6.

$$\theta(x) = t \int g(y)\phi((x - y)t^{1/d})dy ,$$

$$|\theta(x_1) - \theta(x_2)| \leq \int|\phi((x_1 - x_2)t^{1/d} + y) - \phi(y)|dy ,$$

$$\sup_{|x_1-x_2|\leq h} |\theta(x_1) - \theta(x_2)| \leq \omega(ht^{1/d})$$

where ω is the L_1 modulus of continuity of ϕ. Therefore

$$|\theta(x_1) - \theta(x_2)| \leq \eta \quad \text{if } ht^{1/d} \leq \eta' , \quad \text{i.e.,} \quad \text{if } h \leq \eta' t^{-1/d} .$$

We can divide the torus T_ℓ^d into small cubes of size $\eta' t^{-1/d}$, and we will then have $t/(\eta')^d$ cubes. In order to cover the unit ball, we need step functions that are constant on cubes, and an easy estimate provides the bound

$$N \leq [\frac{C}{\rho}]^{t/(\eta')^d} .$$

This almost completes the proof of Lemma 8.6.

Finally we need to show that

$$\inf_G [\lambda(G) + v|G|] = k(v,d) > 0 .$$

If we expand a region by a factor σ, then $\lambda(\sigma G) = (1/\sigma^2)\lambda(G)$ and $|\sigma G| = \sigma^d|G|$. Then

$$\inf_{\sigma>0}\left[\frac{\lambda(G)}{\sigma^2} + v\sigma^d|G|\right] = c(v,d)|G|^{2/(d+2)}[\lambda(G)]^{d/(d+2)}$$

where $c(v,d)$ can be calculated explicitly. Therefore

$$k(v,d) = c(v,d) \inf_{|G|=1} [\lambda(G)]^{d/(d+2)} .$$

A rearrangement argument tells us that the infimum is attained when G is the sphere of unit volume in R^d. This calculates $k(v,d)$ explicitly, and $k(v,d) > 0$. For details see [1]. □

Section 9. The Polaron Problem

A problem that comes up in statistical mechanics, known as the polaron problem, leads to the following question concerning Brownian motion. Does

$$\lim_{t\to\infty} \frac{1}{t} \log E \left\{\exp\left[\alpha \int_0^t \int_0^t \frac{e^{-|s-\sigma|}}{|\beta(s)-\beta(\sigma)|} d\sigma\, ds\right]\right\} = g(\alpha) \tag{9.1}$$

exist, where $\beta(\cdot)$ is the three-dimensional tied-down Brownian motion in the interval $[0,t]$? And how does $g(\alpha)$ behave for large α? A conjecture by Pekar states that

$$\lim_{\alpha\to\infty} \frac{g(\alpha)}{\alpha^2} = \sup_{\substack{\phi\in L_2(R^d)\\ ||\phi||_2=1}} \left[2\iint \frac{\phi^2(x)\phi^2(y)}{|x-y|} dx\, dy - \frac{1}{2}\int|\nabla\phi|^2 dx\right] .$$

We will use our methods to prove the conjecture.

First we note that

$$\begin{aligned}
\int_0^t \int_0^t \frac{e^{-|s-\sigma|}}{|x(\sigma)-x(s)|} ds\, d\sigma &= 2\int_0^t d\sigma \int_0^t \frac{e^{-(s-\sigma)}}{|x(s)-x(\sigma)|} ds \\
&= 2\int_0^t d\sigma \int_\sigma^\infty \frac{e^{-(s-\sigma)}}{|x(s)-x(\sigma)|} ds + o(t) \\
&= 2\int_0^t F(\theta_s\omega) ds + o(t)
\end{aligned}$$

where θ_s is the shift and

$$F(\omega) = \int_0^\infty \frac{e^{-\sigma}}{|x(\sigma)-x(0)|} d\sigma .$$

By our large deviation results, one expects $g(\alpha)$ to exist in (9.1) and to be given by the variational formula

$$g(\alpha) = \sup_Q [2\alpha E^Q F(\omega) - H(Q)]$$

where $H(Q)$ is the entropy relative to Brownian motion of the stationary process Q and Q varies over all stationary processes with values in R^3. There are two technical problems here. The first is the fact that we have tied-down Brownian motion and not free Brownian motion. For t large there is very little difference, and this can be made precise. The details are in [3]. A more serious problem is the fact that Brownian motion does not satisfy the conditions for obtaining upper bounds. The lower bound, hoever, follows painlessly by our methods. To get upper bounds, one notices that if we replace Brownian motion by the Ornstein-Uhlenbeck process with generator

$$\frac{1}{2}\Delta - \varepsilon x\cdot\nabla$$

for some small $\varepsilon > 0$, the theory applies, and moreover the expectation for Brownian motion is dominated by the expectation of the OU process for every $\varepsilon > 0$. If $H_\varepsilon(Q)$ is the entropy relative to the OU process and $H(Q)$ is the entropy relative to Brownian motion, then for any Q

$$H_\varepsilon(Q) = H(Q) + \varepsilon \int ||x(0)||^2 dQ - \frac{3\varepsilon}{2} \geq H(Q) - \frac{3\varepsilon}{2} .$$

Therefore

$$\limsup_{t\to\infty} \frac{1}{t} \log E \{\exp[\alpha \int_0^t \int_0^t \frac{e^{-|\sigma-s|}}{|\beta(\sigma)-\beta(s)|}]\}$$
$$\leq \sup_Q [2\alpha E^Q F(\omega) - H_\varepsilon(Q)] \leq \sup_Q [2\alpha E^Q F(\omega) - H(Q)] + \frac{3\varepsilon}{2} .$$

By letting $\varepsilon \to 0$ we obtain that the limit (15.1) exists and g is given by

$$(9.2) \qquad g(\alpha) = \sup_Q [2\alpha E^Q F(\omega) - H(Q)] .$$

Using (9.2) and Brownian scaling, one can get

$$(9.3) \qquad \frac{g(\alpha)}{\alpha^2} = \sup_Q [\frac{2}{\alpha^2} E^Q\{\int_0^\infty \frac{e^{-t/\alpha^2}}{|x(t)-x(0)|} dt - H(Q)\}] ,$$

and now we have to see what happens to

$$E^Q\{\frac{2}{\alpha^2}\int_0^\infty \frac{e^{-t/\alpha^2}}{|x(t)-x(0)|} dt\} \text{ as } \alpha \to\infty .$$

Writing q(t,dx,dy) for the two-dimensional distribution of x(0) and x(t) under the stationary process Q, we have

$$\lim_{\alpha\to\infty}\frac{1}{\alpha^2}\int_0^\infty e^{-t/\alpha^2}dt\int\int\frac{q(t,dx,dy)}{|x-y|} = \lim_{t\to\infty}\int\int\frac{q(t,dx,dy)}{|x-y|}$$
$$= \int\int\frac{q(dx)q(dy)}{|x-y|}\,.$$

This is not quite correct. However, if Q is ergodic, the independence of x(0), x(t) in an average sense, for $t \to \infty$, is enough to give the final answer. This argument is essentially correct.

There is also a serious problem of interchanging sup and limit on α. If we could carry this out we would have

$$\text{(9.4)}\qquad \lim_{\alpha\to\infty}\frac{g(\alpha)}{\alpha^2} = \sup_Q\,[2\int\int\frac{q(dx)q(dy)}{|x-y|} - H(Q)]$$
$$= \sup_q\,[2\int\int\frac{q(dx)q(dy)}{|x-y|} - I(q)]$$

by the contraction principle. Since

$$I(q) = \frac{1}{8}\int\frac{|\nabla f|^2}{f}\,dx$$

if $q(dx) = f(x)dx$ and $I(q) = \infty$ otherwise, the variational formula in (9.4) reduces to Pekar's conjecture. Incidentally, the unbounded nature of the function $1/|x|$ causes additional technical problems that need to be handled. All this has been rigorously justified, and the details can be found in [3].

Section 10. Large deviations and laws of the iterated logarithm

Let $\ell(t,\cdot)$ be the local time of the one dimension Brownian motion defined by

$$\ell(t,y) = \int_0^t \delta(\beta(s) - y)ds$$

One knows that $\ell(t,y)$ is jointly continuous in t and y. If we define

$$\hat{\ell}(t,y) = \frac{1}{\sqrt{t}}\,\ell(t,\sqrt{t}y)$$

and

$$\tilde{\ell}(t,y) = \frac{1}{\sqrt{t\log\log t}}\,\ell(t,\sqrt{t/\log\log t}\;y)$$

then the distribution of $\hat{\ell}(t,y)$ is independent of t by Brownian scaling. One can get functional laws of the iterated logarithm $\tilde{\ell}(t,\cdot)$ by showing that the set of limit points of $\tilde{\ell}(t,\cdot)$ as $t \to \infty$ are precisely the set of subprobability densities

p(y) with $\int p(y)dy \leq 1$ and $\frac{1}{8} \int \frac{[p'(y)]^2}{p(y)} dy \leq 1$. In particular we can take functionals F which are nice and obtain

$$\limsup_{t\to\infty} F(\tilde{\ell}(t,\cdot)) = \sup_{C} F(p(\cdot))$$

where C is the set of limit points described earlier.

If we take $F(p(\cdot)) = p(0)$ we obtain

$$\limsup_{t\to\infty} \frac{\ell(t)}{\sqrt{t \log \log t}} = \sqrt{2} \text{ a e}$$

If we take $F(p()) = \inf_{q} [\ell : \int_{-\ell} p(y)dy = 1]$ then we obtain (taking liminf rather than limsup)

$$\liminf_{t\to\infty} \sqrt{\frac{\log\log t}{t}} \sup_{0\leq s\leq t} |\beta(s)| = \frac{\pi}{\sqrt{8}} .$$

This is the so-called "other law of iterated logarithm". There are several other examples that one can think of. For some of the details see [2].

Concluding Remarks

The large deviation theory that we have developed during these lectures depends on rather stringent assumptions on the transition probabilities p(t,x,dy). These assumptions are strong enough to ensure the existence of at most one invariant probability measure for the Markov Process. If we were to drop this strong ergodicity assumption then the large deviation rate, even when they exist could start to depend on the starting point. To be more precise, if the Markov Process were to admit several invariant probability measures, the extremals among them being ergodic, then the large deviation rate for the types of sets that we had considered before, namely

$$P_x[L(t,\omega,\cdot) \in A]$$

could have the following type of behavior:

$$P_x[L(t,\omega,\cdot) \in A] = \exp[-t\, I_\alpha(A) + o(t)]$$

for almsot all x ω·r·t α. Here α is an ergodic invariant probability measure and $I_\alpha(\cdot)$ is computed in terms of a rate function depending on α. We can also start with initial distribution α and then

$$P_\alpha[L(t,\omega,\cdot) \in A] = \exp[-t\, \hat{I}_\alpha(A) + o(t)]$$

where $\hat{I}_\alpha(\cdot)$ is computed in terms of a slightly different rate function also depending on α. I_α takes care of large deviations in the evolution where $\tilde{I}_\alpha$ takes care of large deviations in the initial conditions as well. There are some interesting examples of infinite particle systems where computations have been made to illustrate such behavior. The relevant references are [6], [7] and [8].

Bibliographical Remarks:

The results outlined here appeared originally in several articles. There are now several sources available that provide a general survey of the large deviation theory along with a list of references: Varadhan [10], [11], Stroock [9] and Ellis [5] are good sources. The missing details in some of the applications of Sections 6, 8, 9, 10 can be found in [1], [2] and [3] and [4].

References

[1] Donsker, M. D. and Varadhan, S.R.S. The Asymptotics of The Wiener Sausage, Comm. Pure Appl. Math. Vo. XXVIII, 1975, 525-565.

[2] Donsker, M.D. and Varadhan S.R.S. On Laws of Iterated Logarithms for Local Times, Comm. Pure Appl. Math. Vol. XXX, 1977, 707-753.

[3] Donsker, M.D. and Varadhan, S.R.S. Asymptotics for the Polaron, Comm. Pure Appl. Math. Vol. XXXVI, 1983, 505-528.

[4] Donsker, M.D. and Varadhan, S.R.S. Large Deviations for Stationary Gaussian Processes, Comm. Math. Physics 97, 1985, 187-210.

[5] Ellis, Richard S. Entropy, Large Deviations and Statistical Mechanics, Springer-Verlag, New York, 985.

[6] Cox and Griffeath, Large Deviations for some infinite particle system occupation times. Contemp. Math. Vol. 41, 1985, 3-54.

[7] Bramson, Cox and Griffeath. Large deviations for infinite particle systems, preprint, 1985.

[8] Lee, T.Y. Large Deviation theory for the empirical density of non interacting infinite particle systems, 1986, Thesis NYU.

[9] Stroock, D.W. An introduction to the theory of large deviations Springer-Verlag, New York, 1984.

[10] Varadhan, S.R.S. Large Deviations and Applictions, SIAM, Philadelphia, 1984.

[11] Varadhan, S.R.S. Large Deviations and Applications, Expo. Math. 3, 1985, 251-272.